DOCUMENT RESUME

ED 295 587 IR 013 180

TITLE Telecommunications Policy Research Conference. Regulation, Deregulation & Competition Section. Papers.
INSTITUTION Telecommunications Policy Research Conference, Inc., Washington, DC.
PUB DATE Sep 87
NOTE 120p.; Papers presented at the Telecommunications Policy Research Conference (15th, Airlie, VA, September 27-30, 1987). For the conference program and additional papers, see IR 013 160-182.
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC05 Plus Postage.
DESCRIPTORS Case Studies; Comparative Analysis; *Information Networks; Regression (Statistics); *State Agencies; State Legislation; *Telecommunications; *Telephone Communications Industry
IDENTIFIERS *Deregulation; *Pricing

ABSTRACT

Three papers discuss aspects of telecommunications regulation in a deregulated environment. The first paper, "Implementing Telephone Deregulation: The Political Economy of State Regulation in the Post-Divestiture Era" (Paul E. Teske), analyzed the variation in state regulation of local telephone operating companies using regression analyses across the 50 states and comparative case studies of several states. It was found that different regulatory structures and attitudes explain much of the policy choice variation. Data are presented in four tables and a 27-item bibliography is attached. The second paper, "Network Externalities and Critical Mass in New Telecommunications Services" (David Allen), examines information services in France and Japan in order to develop a theory of the "critical mass" of network externalities that play a central role in the demand for new networks. Two appendixes, 17 footnotes, and an 11-item bibliography are included. The third paper, "A Comparison of Rate Base Rate-of-Return and Ceiling Price Regulation" (Neal C. Stolleman), uses comparative static analysis to examine two alternative regulatory systems from the perspective of a multi-output, profit maximizing exchange carrier. (Author/EW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED295587

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

TELECOMMUNICATIONS POLICY RESEARCH CONFERENCE

Regulation, Deregulation & Competition Section

Airlie, Virginia

September 27-30, 1987

IR013180

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Edward Zajec

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

BEST COPY AVAILABLE

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

IMPLEMENTING TELEPHONE DEREGULATION

The Political Economy of State Regulation in the Post-Divestiture Era

by Paul E. Teske

State University of New York, Stony Brook

Prepared for Delivery at the

1987 Telecommunications Policy Research Conference
Airlie House, Virginia
September 30, 1987.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Edward Zajec

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Paul E. Teske

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

IR013180

ABSTRACT

This paper analyzes the variation in state regulation of local telephone operating companies in the five year period from 1982 through 1986. With the exogenous shock of the AT&T divestiture and FCC deregulatory decisions, all state regulators were faced with the same basic choices of changing rate structures and allowing competition or not. State regulatory variation is explained according to political-economic theories of environmental and technological differences, interest group pressure, and institutional differences in regulatory structures. The methodology includes regression analyses across the fifty states and comparative case studies of several states, including Illinois, New York, Nebraska, Vermont, Arizona, California, and New Jersey. The hypothesis that different regulatory structures and attitudes explain much of the policy choice variation is confirmed. The study also finds increasing involvement by state legislatures in this issue, often preempting or forcing the policy decisions of state public utility commissions.

INTRODUCTION

Although states had regulated telephone service for seventy-five years, the exogenous federal shock of the AT&T divestiture in the early 1980s created several new problems for state public utility commissions (PUCs). How should prices be changed and where should competition be allowed? The federal regulators at the Federal Communications Commission (FCC) presented a model, backed by economic research, of reducing toll prices and increasing local end-user access rates, while encouraging competition in almost every conceivable service provided over the telephone network. While the economic efficiencies that could be captured under this approach are appealing to regulators, the policy presented two major problems for them. First, this policy appeared regressive relative to household incomes, which might threaten long-held universal service goals. Second, it would force state regulators to raise prices on a widely-shared, easily identified service that all voters use. Furthermore, eliminating cross-subsidies through the deaveraging of prices generally would create winners and losers, with winners more likely to thank technology than the regulators, and losers able to identify the source of their discontent most readily. It is easily understandable why Congress did not pass a single significant piece of legislation related to telephone issues in the past 15 years.

Some state PUCs have used the FCC model, or their own version, to implement sweeping changes, while others have fought a rear-guard action to preserve the status quo. Why have states varied so much in their policy approaches to telephone regulation? Are they reacting to different environmental economic or technological conditions? Are interest groups more successful in pressuring for different policies across the states? Or, do different institutional arrangements and regulatory attitudes account for the differences?

The methodology utilized here includes regression analyses across all fifty states to isolate the causal factors in a rigorous quantitative sense and case studies of innovative states to identify the reasons for telecommunications regulatory choices.

Since the 1960s, scholars have tried to explain policy differences across states using indices of state liberalism, innovativeness, party competition, income variation, and other characterizations (see, for example, Cnudde, 1969; Dye, 1969; Sharkansky, 1969; Walker, 1969; Jennings, 1979; Wright, et. al., 1985). The explanations have been somewhat unsatisfactory, in part, because the research issues have been too broad. We can get a better handle on state cross-sectional variation by focusing on a policy that faced an abrupt exogenous change rather than a gradual evolution, with similar institutions making the decisions in all states.

ROLE OF STATE REGULATION

Public utility regulation began at the state level eighty years ago. New York, Wisconsin, Illinois, and New Jersey were among the first states pushed by an alliance of progressive politicians and industry interests to establish independent regulatory bodies. Politicians defended these agencies as necessary to prevent the new "home-intruding" natural monopolies such as telephones, electricity, and water service from abusing the public interest. The public utility commissions (PUCs, in some states called public service commissions, commerce commissions, corporation commissions, or public utility boards) evolved to focus on intrastate telephone service while the FCC regulated interstate telephone and telegraph services.

The PUCs are currently fairly diverse in structure and function. All include an odd number of commissioners, ranging from one to seven. Most PUCs have overlapping terms for commissioners, ranging from four year terms in most states, up to ten years. In about three-quarters of the states, the governors appoint the commissioners, and the governor usually selects the Chairperson. In eleven states voters directly elect commissioners, while two other states have the legislature elect commissioners. Increasing professionalism has been established in appointments; several states employ non-partisan nominating committees that prepare a list of qualified candidates from which the governor must choose.

Although all PUCs hold quasi-judicial power, with their decisions subject to judicial review, institutional responsibility varies greatly across states. In some states, such as Virginia, the PUC assumes almost all regulatory functions, including insurance, banking, corporate charters, and professional licensing. In other states, like New Mexico, the PUC performs far more limited regulatory functions, in only a few industries.

Regulatory budgets varied greatly across the states from $1 to $56 million in 1984. Some PUC staffs are explicitly charged with protecting small consumer interests, while other states have their Deputy Attorney General or a Public Advocate play this role.

Utility regulation moved out of the political backwaters in the 1970s when the Arab oil embargo raised energy prices causing electricity rates, nuclear power, and conservation to became significant issues. Telecommunications issues remained fairly simple and straightforward on the state level until the Consent Decree for divestiture in 1982. Divestiture, plus inflationary pressure on local rates in the four years prior to the breakup, changed the level of interest and importance for state telecommunications regulation.

Significant turmoil at the federal level in the 1970s was both a cause and result of Congressional failure to update the 1934 Communications Act, which had established the FCC. In parallel, many state PUC enabling statutes were passed decades ago and have needed legislative up-dating because they did not deal with the issues of competition and deregulation, and hamstrung regulatory options.

The lack of clear federal legislation and the difficulties of separating the telephone network into interstate and intrastate pieces led to major disagreements between the FCC and state PUCs on regulatory jurisdiction (See Maher, 1985 and Noam, 1983). The conflict intensified as the FCC's move toward efficient pricing and competition became evident to state regulators. The FCC won several jurisdictional disputes in court. In 1986, however, the Supreme Court's Louisiana Public Service Commission v. FCC, decision prevented FCC pre-emption of intrastate depreciation practices. The tide seemingly stemmed at that point, and states took more self-confident regulatory steps.

REGULATORY STAKEHOLDERS

Several stakeholders with interests in these issues participate in state regulatory proceedings. The regulated local exchange companies (LECs) have the most at stake. They favor changing prices, with competition delayed until they are unencumbered and allowed to provide all possible services. Potential competitors, especially long-distance carriers, also favor price changes, particularly reductions in carrier access charges, and the right to compete in all markets with minimal encumbrances. Unionized communications workers generally oppose changes in the status quo, particularly since new competitors tend not to be unionized. Cable television companies, since 1984 free of most price regulation, would like to get into data carriage and other interactive telecommunications services.

Several distinct consumer stakeholders can be identified as well. Large telecommunications-using businesses, especially in the service sectors of our economy, favor changing prices and increased competition. New York Telephone discovered that the sectors with the greatest usage, and thus the most to gain, are the financial, insurance and real estate, transportation and utility, business services, and wholesale trade sectors, all staples of the post-industrial service economy. These firms represent the bypass threat as, nationally, some 45-53% of business toll usage is generated from only 1% of business locations (GAO, 1986).

The incidence of price changes for small businesses is more complicated. Most firms under 500 employees will gain from shifting non-traffic sensitive costs from toll to access prices, and from competition, but do not have nearly as much to gain as larger firms (Bell Communications Research, 1984). The improvement of smaller private branch exchanges (PBXs) and the expansion of shared tenant services should eventually increase options for these firms.

Residences have the most to lose from moving prices toward economic costs, particularly rural residences. The median household will probably pay more as end-user prices rise and toll rates fall. Consumer groups argue that the average residential bill rose 20% from 1984 through 1986, although the FCC notes that CPI changes in telephone bills increased more from 1980-1983 than from 1983-1986 (Consumer Federation of America, 1986). On the other hand, enhanced choices through competition, such as telephone instrument purchases, can also help to stabilize or reduce overall telephone bills for residential consumers. Consumer opposition seems strongest against local increases and implementation of local measured service.

QUANTITATIVE ANALYSIS

I will develop the quantitative analysis prior to the case studies, and then use the cases to supplement the regression results. The two approaches complement each other well. The weaknesses of case studies, including possible selection bias and not being able to hold other factors constant when evaluating the influence of a variable, are balanced by the regression approach. The regression approach alone would not be sufficient either, due to the difficulty of obtaining perfect comparative operational measures for the relevant theoretical concepts.

First, the policies undertaken by states must be characterized.

Dependent Variables

The range of choices open to regulators was presented in the case studies. In essence, regulators could move prices towards economic costs rapidly, slowly, not at all, or they could move prices further from economic costs. Economists' prescriptions for efficiency generally include an increase in end-user access charges, local measured service rather than flat-rate pricing, a reduction of intrastate and interstate toll rates, and an increase in depreciation schedules to reflect the more rapid obsolescence of equipment in a competitive environment. The decisions related to the appropriate degree of competition and regulation focus on intraLATA toll, shared tenant services, Centrex, pay telephones, and capacity resale.

The advantages to changing prices include increased economic efficiency and deterrance of uneconomic bypass by large users. The advantages of competition include greater consumer choice, innovation, and reduction in regulatory costs and distortions. Part of the difficulty in implementing these changes is due to the harm they might cause to universal service, the dollar losses many if not most residential consumers will feel, and possible harm to new competitors.

How can we compare the choices states have made about telephone pricing? Each local telephone operating company provides dozens of services, each with different prices. It is nearly impossible to develop a single index for all prices.

An obvious measure of pricing policy is whether or not the general principles of economic efficiency are being followed. A state can be characterized as following efficient pricing if either an explicit end-user access charge is assessed (recognizing that local service rightly consists of two parts, access and usage), or if toll rates are decreased within the state. Fifteen states implemented one or both of these policies from 1982 through 1986, and it is the pricing dependent variable.

Competition choices are easier to characterize. The issue that is most controversial is intraLATA competition and only about one-third of all states have allowed it.

Independent Variables

Several independent variables from the theoretical literature on regulation are tested in this cross-sectional analysis. The factors fall into two broad categories, external and internal explanations. The external factors are either economic and technological (environmental) in nature or they represent interest group strength and activity. The internal factors capture differences in regulatory institutional structures and the attitudes of regulators.

The environmental factors include the 1984 level of National Exchange Carrier Association (NECA) subsidy funds that flow in or out of the state, the percentage of TV households in the state served by cable television, and the average access loop costs in the state. The NECA variable, which ranges from -190 in NewJer sey to +254 in Florida, is used to capture different inter- and intrastate calling patterns and experience with subsidies. The cable TV variable, which ranges from 4% to 70%, is designed to pick-up the extent of penetration and sunk facilities costs by a possible technological competitor. The average access loop costs, which ranged from about $6 per month in D.C. to about $24 per month in Wyoming, according to 1984 FCC figures, reflect how far the state needs to go with end-user charges to price efficiently.

The interest group variables include dummy variables for states served by the most aggressive regional holding company, U.S. West, a measure of concentrated large users, the Fortune Service 450 headquarters, and William Gormley's measures of grassroots and state-funded consumer advocacy in PUC proceedings (Gormley, 1983).

U.S. West is univerally acclaimed as the most aggressive RHC and has also been somewhat unique in pursuing a state legislative strategy to get around PUC regulation. The Fortune 450 variable includes the headquarters of the largest white-collar telecommunications-using firms, including banks, brokerage houses, financial firms, insurance companies and utilities. Gormley's variables are a simple bivariate measure of intensity of advocacy participation in PUC proceedings.

The internal variables include whether or not commissioners are elected or appointed, the regulatory budget available, the legislative party control in the state, and a Wall Street measure of the regulatory climate of the state PUC. The legislative party control variable is 1 if both houses are Republican, 0 if the two houses are under different control, and -1 if both houses are Democratic. The regulatory climate ranges from 1 to 9 as assessed by Regulatory Research Associates.

The regressions are tested for both pricing and competition decisions using all seven variables from the external model and four variables from the internal model. A logit procedure is used since the dependent variables represent a discrete 0,1 choice.

For pricing, this regression is significant at the 98% level with two highly significant variables (95% or higher) and two more above 89% significance. (The auxiliary adjusted R-squared (AAR) values show the results of regressing that independent variable on all others. Only the Fort 450 and budget variables show high AARs, with the result that the Fort 450 variable is insignificant and operates in the wrong direction. In this pricing regression minus the budget variable, Fort 450 is significant at the 95% level and with a positive sign. The regression as a whole, however, is much weaker.).

Logit Regression : For Pricing Decisions

Chi-Square 22.7, 98% significant.

Variable	Theory	Beta	Std. E.	Sgnf.	AAR
Fort 450 HQs	+	-0.09	0.10	66%	.71
Grassroots Advcts.	-	0.82	0.94	62%	.06
Govt-Funded Advcts.	-	0.20	0.88	18%	.05
US West	+	0.94	1.05	63%	.20
NECA	-	-0.02	0.01	89% *	.25
Cable Television	-	-0.04	0.05	60%	.09
Access Loop Costs	-	0.01	0.02	55%	.41
Elected Comm.	-	1.36	1.09	79%	.08
Budget for Reg.	+	0.31	0.14	97% **	.67
Legislative Control	+	1.04	0.58	92% *	.15
Regulatory Climate	+	0.39	0.20	95% **	.13

The highly significant variables are the regulatory climate and the regulatory budget, with the legislative party control and NECA funds nearly as significant. The regression as a whole performs very well, predicting 42 of 49 cases correctly, or 87% (a state is considered to be correct if the model predicted .5 probability or higher and the state made a positive choice, or .49 or lower and the state did not make a positive choices). The state cases predicted incorrectly include Alabama, Delaware, Maryland, Massachusetts, Michigan, New Jersey, and Wyoming. The signs on the non-significant variables are in the expected theoretical directions, with the exception of access loop costs and elected commissioners.

We must consider the strength of these relationships as well as their significance to understand the impact of the independent variables. The following table shows the probability impact on price decisions of moving each variable from its lowest value (LV) to a standard deviation below the mean (SDBM), to a standard deviation above the mean (SDAM), to its highest value (HV), with the other variables evaluated at their means. When all variables are evaluated at their means, the probability is .28.

Probability of Price Change for Variable at

	LV	SDBM	SDAM	HV	StD Change
Fort 450	.41	.41	09	.002	.32 *
Grsrts. Advcts.	.17	.17	.31	.31	.14
Gvt-Fnd Advcts.	.22	.22	.25	.25	.03
US West	.19	.19	.32	.38	.13
NECA	.93	.53	.08	.001	.45 *
CATV	.63	.32	.17	.11	.15
Acs Loop Csts	.11	.17	.32	.53	.15
Elected Comm.	.17	.17	.37	.45	.20
Budget for Reg.	.03	.03	.85	1.00	.82 **
Leg Control	.13	.13	.43	.54	.30 *
Reg Climate	.05	.10	.48	.63	.38 *

Based on a standard deviation departure from the mean, the regulatory budget, the NECA subsidy, the regulatory climate, the Fortune 450 headquarters, and the legislative party control have the largest effects on the probability of a price change decision. For example, with the other independent variables held at their means, a change in the regulatory climate from one standard deviation below its mean to a standard deviation above results in a 38% increase in the likelihood of a price change decision.

The regression for competition decisions is also significant at the highest levels (99%).

Logit Regression : For Competition Decisions

Chi-Square 38.0, 99% significant.

Variable	Theory	Beta	Std. Error	Significance
Fort 450 HQs	+	0.15	0.12	80% *
Grassroots Advcts.	+	-0.49	1.44	26%
Govt-Funded Advcts.	+	5.63	2.74	96% **
US West	+	8.40	3.92	97% **
NECA	-	-0.01	0.01	54%
Cable Television	+	0.06	0.07	62%
Access Loop Costs	-	-0.01	0.03	34%
Elected Comm.	-	-4.56	3.05	87% *
Budget for Reg.	+	-0.08	0.10	55%
Legislative Control	-	-3.53	1.49	98% **
Regulatory Climate	+	0.55	0.34	89% *

Several variables are significant above 95% levels, with the expected sign, including legislative party control, U.S. West states, and governmentally funded advocacy. The regulatory climate, elected commissioners and Fort 450 headquarters are in the 80-90% significance range. All signs are in the expected theoretical direction, except for two insignificant variables, grassroots advocacy and the regulatory budget.

The model predicts 44 of the 49 cases correctly, or 90%. The states predicted inaccurately include California, Kentucky, New Jersey, Ohio, and Pennsylvania.

The next table shows the strength of the relationships. With all variables evaluated at their mean value, the model predicts a .09 probability of a pro-competition choice.

Probability of Competition Choice for Variable at

	LV	SDBM	SDAM	HV	Std Dev Change	
Fort 450	.02	.02	.41	.99	.39	*
Grsrts. Advcts.	.11	.11	.07	.07	.04	
Gvt-Fnd Advcts.	.01	.01	.60	.60	.59	**
US West	.01	.01	.82	.97	.81	**
NECA	.38	.15	.05	.01	.10	
CATV	.01	.05	.15	.28	.10	
Acs Loop Csts	.19	.13	.06	.03	.07	
Elected Comm.	.26	.26	.01	.01	.25	
Budget for Reg.	.16	.16	.04	.002	.12	
Leg Control	.56	.56	.01	.001	.55	**
Reg Climate	.01	.02	.31	.51	.29	

Based on a standard deviation departure from the mean value, the U.S. West companies, governmentally-funded advocates, legislative party control, and Fortune 450 headquarters have the strongest impact on competition policy choices. With other variables held at their means, a U.S. West state has a 96% increase in its likelihood of allowing intraLATA competition over any other state.

Discussion

The strongest and most significant variables for pricing include the budget, NECA subsidies, regulatory climate, Fortune 450 headquarters, and legislative control. For competition choices, the key factors are U.S. West, governmentally-funded advocates, legislative control, Fortune 450 and regulatory climate. Only legislative control, regulatory climate and Fortune 450 are highly important in both decisions. Clearly, although price and competition policies are linked, some states do not look at them as such, at least in the short-run.

The results suggest that interest groups, particularly U.S. West, large users and government consumer advocates, affect regulatory decisions, especially competition decisions. For price changes, bureaucratic factors such as the regulatory budget and regulatory climate are more important. The regulatory climate also influences competition choices. The influence of the state legislature looms in both.

CASE STUDIES

The case studies show the range of policy choices available to regulators, and explain why regulators in a given state choose their specific policy. The policies will be summarized and the politics examined in Illinois, New York, Nebraska, Vermont, Arizona, California, and New Jersey. This order roughly reflects a continuum from more to less deregulatory policies.

The politics of state telephone regulation have led to several interesting and different decisions. The Illinois PSC moved aggressively to eliminate almost all forms of cross-subsidies and to transfer access costs to end users. The New York PSC started changing prices well before divestiture and has moved to eliminate many cross-subsidies, although not as aggressively as in Illinois. The Nebraska legislature approved a bill that would greatly decrease PSC control over rates. The Vermont legislature approved a form of "Social Contract" that will keep local rates down while allowing substantial freedom for the local telephone company in the more competitive services. Arizona voters rejected a referendum that would have forced further telephone deregulation. California legislators instituted a Lifeline program for low-income subscribers while the PUC has maintained several cross-subsidies. New Jersey has competed with California for the lowest local rates in the country, has not allowed intraLATA competition, and has had no legislative action.

Illinois

A tradition of progressive regulation in Illinois has helped lead to many changes in telecommunications regulation. Illinois Bell, the main regulated company in the state, was considered to be one of the best BOCs under the old AT&T structure, and leadership in the company usually led to a high position in AT&T, with the last three AT&T Chairmen having previously been President of Illinois Bell.

Since divestiture the Illinois Commerce Commission (ICC) has moved towards efficient economic pricing more rapidly than any other state in the country. The ICC was the first to unbundle access from local usage and to de-average access rates across the state. The ICC led the way in restructuring local rates by reducing 93 possible plans to two, residential and business, both of which reflected efficiency by implementing off-peak reductions and volume discounts. The ICC also was a leader in accelerating depreciation to eliminate the reserve deficiency, in reducing the business access subsidy to residential consumers, in allowing intraLATA competition in 1987 and intraexchange competition in 1989, in phasing out the intrastate pooling subsidy to rural companies, and in de-averaging LEC toll rates in the state.

The Illinois Commerce Commission started deregulatory activity prior to divestiture by negotiating to establish Metropolitan Service Areas (MSAs) rather than LATAs in the state. As a result there are more local calling areas than in any other state, leading to more interMSA calling and larger competitive zones. This approach tried to encourage competition and to deal with the issue of several small independent companies serving the state outside of Chicago. Since no other state took this approach, it shows the pro-active approach of dealing with divestiture-related issues, even before the actual breakup.

Telephone deregulation in Illinois received a strong push toward a comprehensive approach due to the state's 1983 Sunset Law that would have required the original (1921) public utility regulation legislation (and subsequent amendments) to sunset in 1985. Legislators were forced to deal with ICC issues in a non-incremental fashion and the result for telephones was the "Universal Telephone Service Protection Law of 1985", which itself will sunset in 1991. As such legislation goes across the country (there were 15 laws by 1987), the legislative intent is quite clear and less discretion is given to the ICC. Case-by-case criteria are established for the ICC to judge, but the burden of proof on telephone companies is small.

Why has Illinois moved so rapidly to eliminate cross-subsidies and how have they gotten away with it politically? It appears that the combination of a policy entrepreneur, Phil O'Connor, the ICC chair, and an economics-oriented staff (rated high by Wall Street), in a tradition of economic efficiency, allowed the ICC to work with the legislative Sunset Act to make radical changes. Consumer opposition was blunted by a preponderance of attention to a major nuclear power plant issue in the state. Rural interests were not able to halt de-averaging, as they have in most other states. The rural members of the legislature allowed the subsidy to end, in part because they were assuaged by a 1983 AT&T refund that was distributed disproportionately to rural areas.

The ICC showed considerable concern about bypass and the consequent need to change prices to reflect costs. A Touche-Ross study showed that 29% of the 110 largest business customers of Illinois Bell were bypassing in 1983, with another 31% considering that option.

Phil O'Connor aligned himself with key staff members to push regulatory changes. O'Connor was a Ph.D. political scientist, who, as Governor Thompson's campaign director, became a strong advocate of efficiency and competition in all public utility services, including electricity. O'Connor became a national advocate of deregulation.

The ICC has had a Ph.D economist on the commission over this entire period, who aligned himself with the deregulatory agenda. There were two such economists, the original is now a federal energy regulator. The economic expertise was "strong up the middle", like a baseball team, with the ICC Chair and his staff. The Executive Director was a University of Chicago trained economist with federal regulatory experience. The staff included several economists and policy analysts, who were not tentative in pushing their own agenda. Thus, a policy entrepreneur in Phil O'Connor and an aggressive staff seem to have been the crucial elements in Illinois.

New York

New York is the second most populous state in the nation. In many ways, it is divided between the largest metropolitan area in the country, New York City and its suburbs, downstate, and a more rural, although still quite populous in total, upstate region. All issues in state politics reflect this divide, which also tends to follow a Democratic and Republican party affiliation, respectively.

New York City has an extremely large concentration of major telecommunications-using businesses, particularly large financial firms. Several large firms had already established private microwave communication facilities prior to the AT&T divestiture, so much so that the ensuing congestion created the need for an underground fiber optic link to a Teleport on Staten Island.

The New York Public Service Commission was one of the first established in the country in 1907. Early regulators in New York were influential in national affairs. The PSC evolved to include seven commissioners, no more than four representing a single party, appointed to six year overlapping terms by the Governor, with Senate approval.

In recent years a major influence on the New York PSC was the Chairmanship of Alfred Kahn from 1974-1977, the man who became most associated with industry deregulation (McCraw, 1984). While Kahn focused his attention on energy issues during the post-Arab oil embargo crises, his development of staff interest in and emphasis on marginal cost pricing and sophisticated industrial organization economics led to changes in telecommunications (Anderson, 1981). In fact, the New York PSC moved to change the structure of telephone prices in the late 1970s, without the threat of bypass but for efficiency reasons (Dugan and Stannard, 1985). "Chairman Swidler had been accustomed to fully distributed cost methodologies through a long and distinguished career, principally in the energy field. By contrast, Chairman Kahn was an active proponent of marginal cost pricing" (1). As a result of Kahn's efforts, in both 1977 and 1979, the PSC reduced toll rates by 5% while increasing overall revenues.

For the entire post-divestiture period, the New York PSC was chaired by Paul Gioia, a Democrat and counsel to ex-Governor Carey, who appointed him in 1981. Gioia ran into several conflicts with Governor Cuomo, who took office in 1982, particularly over cost responsibility for the Shoreham nuclear plant. The PSC and its staff of 641 people received an 8 out of 9 in Regulatory Research Associates ratings and was rated above average by Merrill Lynch.

Traditionally active consumer interests are represented in New York by the state-funded Consumer Protection Board (CPB), headed by Richard Kessel with Governor Cuomo's strong support. Although the Shoreham nuclear power plant on Long Island has received the greatest attention from the CPB in recent years, the CPB has participated actively in all major telecommunications proceedings since divestiture.

The Consent Decree in 1982 found the New York PSC dealing with pricing issues. Before the actual divestiture, the PSC reduced intrastate toll rates, continuing the pattern from the 1970s rate cases under Kahn.

The New York PSC continued to use "generic cases" in the post-divestiture period. These cases are not focused on a rate increase hearing, are not limited in time, and can require extensive data gathering by participants. Alfred Kahn used them to examine rate design and marginal cost pricing in his tenure at the PSC, outside of the more politicized context of a rate case.

In 1983 the PSC decided not to prohibit intraLATA competition but required that carriers pay access charges as with interLATA competition. To make the competition fair, however, New York Telephone must assess itself the same charges.

In December 1983 the PSC initiated an examination of the "bypass" threat. In 1985 the PSC ordered reductions in carrier access charges and intraLATA toll rates to reduce the economic incentives to bypass. Local rates were raised, an average of 3%, while intraLATA and carrier access charges were reduced by 14-27%. A lifeline service was offered in this proceeding as well.

Thus, Kahn's pre-divestiture influence and the presence of many concentrated large users caused price changes, which fall short only of those in Illinois. Since Kahn's tenure, the PSC is not as totally committed to marginal cost pricing but it does tend to follow the the major principles.

Nebraska

Nebraska is a low population state with no particular tradition of innovation in telecommunications. Yet, in 1986, the legislature passed and the Governor signed a law to allow radical price deregulation of all services, including local service, which represented the most extreme deregulation in the nation.

Nebraska Bell is part of the U.S. West holding company, universally acclaimed as the most aggressive RHC in the country in pushing for reductions of regulatory constraints. A strong legislative push marks U.S. West's strategy and it has succeeded, with 9 of 15 deregulatory bills in the nation passed among the 14 states served by U.S. West subsidiaries. Jack Mac Allister, U.S. West chairman, called critics of such deregulation, "regulators who don't want to walk away from their jobs" (2).

The legislation, LB 835, was written by U.S. West. It allows, on 90 days notice, increases of up to 10% annually in local prices, unless 2% of affected consumers sign a petition opposing the increase, in which case the PUC must hold hearings and make a decision. Services other than the local exchange are nearly free of rate-of-return regulation. After 1992, telephone companies in Nebraska are totally free to set rates as the market will bear.

The law was passed over the objections of the PSC regulators. The state Attorney General is suing to have the law declared unconstitutional. Three Senate bills have attempted to reverse the deregulatory legislation. PSC President Harold Simpson says, "How do you force good service if you don't have a handle on rates?" (3).

The PSC resisted intraLATA competition originally, and it was delayed in LB 835 until 1989, but the PSC approved competitive petitions by MCI, U.S. Sprint, and AT&T as of July 1, 1987, under the old statute that required a showing of convenience and necessity.

Nebraska has the only unicameral, officially non-partisan elected legislature in the nation. The need to get the legislation through only one house may have aided passage of the bill. State senator John DeCamp was the main sponsor and advocate of the bill. He was voted out of office in November 1986 (with 62% of the electorate voting for his opponent), at least in part due to the unpopularity of the telecommunications law he sponsored.

Former Democratic Governor Robert Kerrey strongly supported the bill, and as lead telecommunications spokesperson for the National Governors Association, sold it as a law that would markedly enhance Nebraska's position in attracting high-technology, telecommunications-dependent firms, as part of an economic development strategy. Kerrey cites as an example a telemarketing firm that handled over one million calls for Kodak when it stopped producing its instant camera, which recently established operations in Nebraska.

The following remarks are indicative of former Governor Kerrey's stand on this issue. "You had US West making the case for this all over its region, and here in Nebraska they had a receptive audience". "If they don't have to worry about some arbitrary commission of Nebraskans setting their prices, they'll come in here and show us what all their technology can do for our people". "If you live in a rural, isolated state like Nebraska, you absolutely need to be connected to the rest of the country. And there is technology coming along that can connect us much more closely. But to get it, we have to move away from arguing, 'What should the price of the product be?' and into 'What should the product be?'"(4). Kerrey's replacement, Republican Governor Kay Orr, recently announced a proposed bill, LB 775, to offer telecommunications and data processing firms, among other businesses, substantial tax credits for creating jobs and expanding in Nebraska.

An irony first pointed out by Gene Kimmelman of the Consumers Federation of American, is that Nebraska's "first mover" advantage may protect them since U.S. West seeks to sell this deregulatory concept to other states and it must show that political and economic harm is minimal (5).

Clearly deregulation in Nebraska was forced by legislators pushing U.S. West's interests and by economic development concerns of governors of both parties.

Vermont

The most prominent element of telecommunications regulation in Vermont is Public Service Board Chairperson Louise McCarren's "Social Contract", unveiled in June 1985. McCarren touted the benefits of the Social Contract, convinced the Vermont Public Advocate, and, finally, the legislation putting this concept into action was approved and signed by Governor Kunin in 1987. The Social Contract allows the Vermont Public Advocate, a member of the Department of Public Service, to negotiate with New England Telephone a five-year contract which is subject to approval according to "public interest" standards by the Public Service Board.

The benefits of the Social Contract to the public include stabilization of local rates, guaranteed minimum levels of plant modernization, guaranteed minimum service quality standards, and non-discriminatory access to long-distance carriers. The advantages to New England Telephone include relaxed rate-of-return regulations, relaxed prior notice regulations, relaxed toll price regulations, and immediately effective tariffs (in contrast to the current 90 days).

Prior to this legislation, Vermont had recognized efficient economic prices and had allowed substantial price changes to move local access prices closer to costs, which was necessary, in part due to Vermont's average access loop costs being the eleventh highest in the nation. In the first post-divestiture rate case, local rates rose over 20% and two-thirds of the total increase was allocated to local rates. In the next rate case, about three-quarters of the increase was allocated to local rates, which rose up to 35%. Thus, the local flat rate rose from $11 per month to $18 per month in Burlington in less than three years, supporting the subsequent push to stabilize local rates. Economic efficiency was further recognized in Vermont regulatory decisions by the institution of intraLATA toll volume discounts and by the expansion of local measured service.

In February 1986 Vermont became the first single-LATA state to approve intraLATA competition for facilities-based carriers and resellers. The decision noted that local operating companies hold non-exclusive franchises, which means that local competition is not precluded, although no competitors have yet emerged.

Thus, deregulation as Social Contract in Vermont was brought about largely by the policy entrepreneurship of Louise McCarren. Her approach was quite different from many other rural states that are fighting a rear-guard action maintaining cross-subsidies.

Arizona

A referendum in Arizona on November 4, 1986, Proposition 100, represented the first chance the voters had to make a direct statement about state level deregulation of telephone service. Referendum instructions advised voters that a "yes" vote "would authorize the Corporation Commission (ACC) to reduce regulation of some telecommunications services, and assure statewide availability and affordability of telephone services using definitions and procedures prescribed by legislation". Voting "no" would mean "keeping the requirement that the Corporation Commission regulate all telephone and telegraph services" (Megdal, 1986). The reason that a referendum was necessary is that the Corporation Commission's (ACC) power is deliniated in the state constitution, which is not true in most states. Proposition 100 effectively would have passed much regulatory authority from the ACC to the legislature. It was defeated by a 54-46% margin.

Mountain Bell, the U.S. West subsidiary, and AT&T spent close to $3 million, through a political action committee, to support Proposition 100. Those opposing it spent approximately $50,000, and thus were outspend by a 60:1 ratio. Mountain Bell is the main regulated local telephone firm in the state, with 1 million customers in the largest cities, while the other nine firms in the state serve about 95,000 people in rural areas. U.S. West was so interested in the referendum because, while Arizona totals only 12% of the RHC's revenues, it represents some 38% of recent growth. Rural telephone companies and rural interests did not oppose the referendum and companion legislation because it included a Universal Service Fund portion.

Reading the meaning of this referendum vote is difficult. Mountain Bell suggests that it shows "customers expressing confidence in the Corporation Commission's regulation of telecommunications services" (6). Another possibility is that voters oppose further deregulation with the sense that deregulation so far has hurt them.

The referendum became the preferred solution when conflicting and unclear legislation was passed. The Arizona House previously passed a Mountain Bell-sponsored bill to implement "flash-cut" deregulation leaving only the local loop for Corporation Commission regulation, which was almost as radical a bill as passed in Nebraska. The House bill was fought hard by the state Residential Utility Consumer Office along with the Communications Workers of America, the ACC, and competitors, but was approved by a 2:1 margin. The Senate sought to achieve slower deregulation, with the ACC as referee, although legislatively prescribed. The Senate bill provided for service by service deregulation, a far less radical approach. The element of legislative control was crucial to the Senate bill. While ACC interpretation of rules would continue to be necessary, they were not longer to be the main regulators.

A previous statute, passed in 1984 by both Houses, had given the ACC the right to deregulate competitive services and to allow multiple providers of a given service. However, the exact legal scope for deregulation was in dispute, so that the vote on Proposition 100 was necessitated. Approval of the referendum would have included approval of the Senate deregulation bill, which voters did not seem to understand. Despite the huge disparity in campaign spending in favor of a bill that we might have expected voters to not be attentive to, voters defeated the proposition.

California

In California, the most populous state, the cross-subsidies in telephone prices have been more extreme than in most of the rest of the country. Therefore, proposals that involve changing rate structures are difficult to implement due to the need to move farther and the large number of stakeholders. California has been a leader in challenging FCC assumptions about competition and network technologies.

California has been innovative in its Lifeline program for low-income households, which was established by legislation in 1984. The 1987 income limits cover households up to $12,100 for 1-2 persons, $14,100 for three person households, and $2,900 for each additional person. The program reduces costs to one-half of the local measured rate or the flat-rate in the jurisdiction.

California has also been innovative in dealing with the relationships between the regional holding company, with its unregulated subsidiaries, and the regulated local telephone company. Its Public Staff prepared the first full audit of the RHC-affiliate relationship, which found that cross-subsidies were occuring and a $4 million withholding of revenue requirement was levied as a penalty. The staff has proposed placing a 5% "royalty" tax on the profits of diversified services of Pac Telesis, the RHC, on the theory that the marketable elements of the company name, goodwill and technological innovation were generated by the telephone side of the business.

Substantial rate restructuring has not been mandated in California. Local rates remain at or near the lowest in the nation. The PUC staff is carefully monitoring bypass and has not seen an acceleration of the process since the pre-divestiture days. The long-term CPUC plan is to shift more non-traffic sensitive costs to end-users over time, but there is no political support to do it quickly.

In a 1984 Generic Competition Proceeding, interLATA toll was viewed as competitive, with AT&T as the dominant carrier. InterLATA toll costs are being reduced by shifting from the Subscriber Plant Factor (SPF) to the Subscriber Line Usage (SLU) allocator for carrier access charges by 1992. The PUC eliminated an 18% surcharge on interLATA rates in 1984.

IntraLATA toll was considered a regulated monopoly and competition was prohibited, except for high-speed private line services. IntraLATA competition was limited to private line services because of the contribution to local service rates, the asymetry that Pacific Bell is prohibited from the interLATA market, and the irreversability of a positive decision. The staff has recommended the SPF to SLU conversion for intraLATA toll as well as interLATA.

The CPUC staff has performed and commissioned several major studies of the network itself, focusing on the questions of whether digital switches and other advanced technology benefit plain old telephone service (POTS), and whether or not ONA and CEI can provide the competitive incentives necessary to deregulate further. The staff has found that equipment is replaced too rapidly and has recommended against accelerating depreciation schedules to reduce the supposed reserve deficiency.

The former President of the PUC, Donald Vial, became a national spokesperson against moving too rapidly on deregulation. He has reservations about the "Social Contract". Since the CPUC staff finds 80-90% of local network services to be monopolistic, Vial says, "We're not going to be moved into deregulation of something that is monopolistic". Vial notes that with intraLATA competition, "The interest is in cherry picking on those toll routes that are most lucrative within the LATA". He views bypass with little concern. "Anytime telco wants more security in its revenue requirements, it says 'bypass' and we're supposed to lie down and do what it wants. Well, we haven't done that". Vial supports royalty payments as a possible answer to the subsidiary separation problem. (7). It is clear that Vial's attitudes were quite different from those of other telephone policy entrepreneurs, like O'Connor, McCarren and Kerrey.

New Jersey

New Jersey is the most densely populated state in the nation. It is a largely suburban state with New York City dominating the northern area's economy and Philadelphia dominating the southern portion. Thus, Benjamin Franklin aptly termed the state "a barrel with taps at both ends". New Jersey has several large businesses but they are not extremely concentrated geographically. New Jersey has been the site of several firsts in telecommunications technology, including the first electronic switch, in part because Bell Labs have been located in the state. Currently, New Jersey is a hotbed for the industry, with AT&T Labs, Bell Communications Research, AT&T International headquarters, and the National Exchange Carriers Association located there.

In 1911, the New Jersey legislature established one of the first fully-empowered regulatory bodies in the nation, the Board of Public Utilities (BPU). The BPU now consists of three commissioners, no more than two from the same party, who serve overlapping six year terms as appointed by the Governor and confirmed by the Senate.

The New Jersey BPU has not had an entrepreneurial figure in recent years, as several of the other state have had. The commissioners over the relevant period have been President Barbara Curran, a Republican lawyer who served in the state Assembly and has been at the BPU since 1980; George Barbour, a Democratic lawyer with Assembly experience who has served at the BPU since 1976 and is President of the NARUC in 1986; and Robert Guido, a Republican actuary appointed in 1985. The BPU and its staff of 323 people received a 5 of 9 from the RRA ratings and average from Merrill Lynch.

In New Jersey small consumers are represented by the Public Advocates Office, the first established in the U.S. The Public Advocate relies on the Deputy Attorney General's office for staff and legal support, and has been active in telephone rate cases.

The New Jersey BPU has increased intrastate toll rates rather than raising local rates, has not allowed intraLATA competition and does not appear to be close to doing so in the near future, has not unbundled access from usage, and has not de-averaged any rates. No intraLATA competition will be allowed until the BPU is convinced that basic rates or universal service are not threatened. The BPU has dealt with Centrex, private lines, coin phones, and other issues in a pragmatic way, allowing more competition and local operating company flexibility when pushed by market considerations.

The average New Jersey residential flat rate rose only 47% from 1972 to 1984, while the CPI grew 151% over that period. Toll rates were not reduced despite New Jersey Bell cost studies showing prices to be five times incremental costs and twice embedded costs.

The BPU has not held major generic proceedings as in several other major states. It has reacted to requests by New Jersey Bell rather than initiating action. Most recently, New Jersey Bell proposed a three year local rate increase moratorium in exchange for reduced regulation on more competitive services, associated with the reduction in federal tax liability. The BPU accepted this arrangement, which was worked out by New Jersey Bell, the Public Advocate, and the BPU staff.

Large business interests have not participated actively in rate proceedings, despite the fact that the communications industry is very strong in NJ. The level of conflict has not been high compared to other states. The major decisions have all been made in the context of rate cases, some of which have been stipulated by the major parties. No state legislative action has taken place, and no entreprenuers have emerged, despite Commissioner Barbour's role as NARUC President in 1986.

Comparison of Cases

The case studies point out the crucial role of political entrepreneurs in formulating telecommunications policies. The important role of the PUC staff analysis is also underscored.

While the presence and role of large users is relevant, it clearly is not the pivotal factor. Regulators in New York and Illinois appear to be responding to this constituency but regulators in California and New Jersey, states with many large users, are not. Furthermore, the Nebraska and Vermont legislatures made innovative regulatory changes without being forced by large users.

Another result in the case studies is the increasing role of legislatures in state telecommunications regulation. U.S. West has used a legislative approach to get around PUCs not favorable to their interests. In other cases legislative action has been spurred by the need to provide enabling legislation for deregulatory action, or to update relevant statutes that were expiring.

Voters appear disenchanted with the early results of telephone deregulation. In addition to the defeat of the Arizona referendum in November 1986, ballot propositions pushing local measured service were defeated handily in Maine and Oregon. Telecommunications deregulation appeared to play some role in the defeat of state Senator DeCamp in Nebraska and even a role, if minor, in the Oregon gubenatorial race.

CONCLUSIONS

Several conclusions can be drawn by combining the results of the case studies and the quantitative analysis. State telecommunications regulations is no longer only the province of PUCs; state legislatures are increasingly involved. Roger Noll's prediction that the rural subsidies would gradually be discovered and not protected now that state legislatures are no longer overly dominated by rural elements (May 1986 American Economic Review), appears only partially accurate. Legislators have not intervened to protect small or rural consumers but have acted to force or encourage deregulation, despite the potential political harm of access cost shifts to local residential rates. In some states, however, continuing rural cross-subsidies have been necessary to get bills approved. The quantitative results support the conclusions that Democraticly controlled legislatures are less likely to support price changes but more likely to encourage competition, with Republican legislatures pushing for the opposite policies.

This state legislative activity is in contrast to Congress not passing a Communications Act re-write in the late 1970s and passing no legislation relating to divestiture and FCC telephone deregulatory actions since that time. Admittedly, the state deregulatory bills are often very complicated and occasionally ambiguous in their effect. The June 1987 AT&T deregulatory bill passed in Texas is a good example of a compromise bill whose effects are uncertain. Some legislation is also clearly pro-consumer, as in California, where legislation established a generous Lifeline program.

Generally, since legislation has not focused on protecting small consumer interests, state PUCs and state-funded advocates must be seen as doing an adequate job in this area. Legislation has instead empowered or forced PUCs to open up competitive segments of the telecommunications industry. In addition to this need to update seventy or eighty year old PUC enabling statutes, economic development and industry competitiveness issues have helped to spur legislative action.

Second, a trend towards regulatory flexibility and a willingness to try competition are evident in all states, albeit to varying degrees. Reduced entry and filing requirements, more rapid tariff changes, banded rates, and the release of services from regulation are among the methods used. In contrast, many state PUCs are still very stubborn about increasing local rates to economic costs. While Illinois, New York, Vermont, and Nebraska have raised local rates, New Jersey and California have resisted. Furthermore, while about one-third of states have allowed facilities-based intraLATA competition by 1987, the intraLATA playing field was not made completely level in most of those states. Only five states have eliminated intraLATA access charges or forced the BOCs to impute the access charges that competitors must pay.

Third, policy entreprenuers have played significant roles on the state level. The quantitative analysis supports the general conclusion of th case studies that regulatory attitudes and policy entrepreneurs are more significant and important factors than environmental variables or interest group pressure. Phil O'Connor in Illinois and Louise McCarren in Vermont are the best examples of PUC Chairpersons aggressively shaping a new agenda in telecommunications regulation. In Nebraska, Governor Kerrey took an active role in getting deregulatory legislation passed, with an emphasis on economic development. In California, Donald Vial fought to oppose price changes. The quantitative results suggest that regulatory attitudes and resources are important correlates of policy. Since individual regulators influence policy outcomes so much, more care should be taken with appointment and elections. Non-partisan nominating list committees are an excellent way to implement this approach for appointed commissioners. Since regulatory budgets and staffs appear to play an important role as well, more thought should be given to the size and quality of the regulatory bureaucracies.

FOOTNOTES

1 "Barriers to Marginal Cost Pricing in Regulated Telecommunications", Dennis Dugan and Richard Stannard, Public Utilities Fortnightly, November 28, 1985, p. 45.

2 "Many States Deregulate Telephone Rates, Hurting Residential Users in Short Run", Bob Davis, Wall Street Journal, September 19, 1986.

3 Ibid.

4 "Phone Deregulation, Phase 2", T.R. Reid, Washington Post, May 27, 1986, A-1.

5 "Nebraska Law Deregulating Phone Rates May Spur Some Changes in Other States", Johnnie Roberts, Wall Street Journal, May 5, 1986.

6 "Telcos Have Bad Election Day; Lose Deregulation, Mandatory LMS Referendums, Key Friend", Telecommunications Reports, November 10, 1986, p. 10-12.

7 "Deregulation Dissent: Vial Urges Caution", Susan Ibis, Telephony Magazine, March 30, 1987, p. 53.

BIBLIOGRAPHY

Anderson, Douglas. 1981.
Regulatory Politics and Electric Utilities.
Auburn House, Boston.

Bell Communications Research. November 1984. "The Impact of End-User Charges on Small Business".

California PUC Policy and Planning Division. May 1987. "Competition in Local Telecommunications: A Report to the Legislature".

Connecticut Department of Public Utility Control. December 1986. "DPUC Investigation into Competition for Intrastate Interexchange Telecommunications Service".

Cooper, Mark and Kimmelman, Gene. December 1986. "Divestiture Plus Three: Still Crazy After All these Years". Consumer Federation of America.

Cnudde, Charles F. and Donald J. McCrone. 1969. "Party Competition and Welfare Policies in the American States", American Political Science Review, 63:858-66.

Davis, Bob. 1986. "Many States Deregulate Telephone Rates, Hurting Residential Users in Short Run", Wall Street Journal, September 19, 1986.

Dye, Thomas R. 1969. "Income Inequality and American State Politics", American Political Science Review, 63: 157-62.

GAO Report. August 1986. "Telephone Communications: Bypass of the Local Telephone Companies".

Gormley, William T. 1983.
The Politics of Public Utility Regulation.
University of Pittsburgh Press.

Jennings, Edward T, Jr. 1979. "Competition, Constituencies, and Welfare Policies in American States", American Political Science Review, 73: 414-29.

Kahn, Alfred E. 1984. "The Next Steps in Telecommunications Regulation and Research." Public Utilities Fortnightly, July 19, 1984, 13-18.

Maher, William F. Jr. 1985. "Legal Aspects of State and Federal Regulatory Jurisdiction Over the Telephone Industry: A Survey." Center for Information Policy Research Monograph, Harvard University.

McCraw, Thomas, K. 1984. Prophets of Regulation. Harvard University Press, Cambridge, Mass.

Megdal, Sharon Bernstein. June 1986. "The Political Economy of Telecommunications Deregulation in Arizona." Draft prepared for Forum on Telecommunications Deregulation, Tucson.

Nadel, Mark. 1986. "The Changing Mission of Telecommunications Regulators At the State Level", Aspen Institute Conference, August 26-30, 1986, Aspen, Colorado.

National Telecommunications and Information Administration Office of Policy Analysis and Development. October 1986. "Telephone Competition and Deregulation: A Survey of the States".

Noam, Eli M. December 1983. "Federal and State Roles in Telecommunications: The Effects of Deregulation." Research Working Paper, Columbia Research Program in Telecommunications and Information Policy.

Noll, Roger G. 1985. "State Regulatory Responses to Competition and Divestiture in the Telecommunications Industry." American Enterprise Institute Public Policy Week.

Reid, T.R. 1986. "Phone Deregulation, Phase 2", Washington Post, May 27, 1986, A-1.

Roberts, Johnnie. 1986. "Nebraska Law Deregulating Phone Rates May Spur Some Changes in Other States", Wall Street Journal, May 5, 1986.

Sharkansky, Ira and Richard Hofferbert. 1969. "Dimensions of State Politics, Economics and Public Policy". American Political Science Review, 63: 867-79.

Telecommunications Reports. 1986-June 1987. Capital Publishing Inc.

Telematics,
The National Journal of Communications Business and Regulation, monthly January-May 1987. Prentice Hall Law and Business.

Ubis, Susan. 1987. "Deregulation Dissent: Vial Urges Caution." Telephony Magazine,
March 30, 1987, p. 52-57.

Walker, Jack. 1969. "The Diffusion of Innovation Among American States", American Political Science Review. September: 880-99.

Wright, Gerald, Robert Erikson and John McIver. 1985. "Measuring State Partisanship and Ideology with Survey Data". Journal of Politics, May: 469-89.

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Network Externalities and Critical Mass in New Telecommunications Services

David Allen
Massachusetts Institute of Technology

Paper Presented to the
Fifteenth Telecommunications Policy
Research Conference

Airlie House, Virginia
September 27-30, 1987

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Edward Zajec

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

IR013180

ABSTRACT

Network externalities - the obvious, though implicit requirement that there be a group of subscribers if communications are to occur - play, it appears, a central role in the demand for new networks. Concern in the US today focuses on the conditions necessary for new information services; these services constitute, at least in some respects, new networks. The development of ISDN, a more worldwide and longer term subject of regulatory attention, is also a case of new networks - and in this case, network "newness" is even more fundamental and thorough-going.

In an effort to describe the demand for new networks, this paper investigates the critical mass phenomena that characterize network externalities. Experience with the Minitel information service in France serves as an empirical base for the analysis. The paper then draws implications for the domain appropriate to regulation and for universal service pricing under liberalization. It concludes with a speculation about the role of mixed economy, illustrated by the construction of Japanese universal service.

In addition to Charles Jonscher and Jeff Rohlfs, who are cited in the text of the paper, I also quite appreciate discussions with Cristiano Antonelli, Henry Elkington, Alain de Fontenay and Paul Verdin. Hajime Oniki, Greg Duncan and Haruko Yamashita will also find that their observations have shaped this text. While I did not always convince them of the arguments here, their generous engagement with the issues has rescued me from pitfalls and allowed me to leaven the presentation with a measure of some added grace. The gaffes that remain, alas, I must answer for.

Network Externalities and Critical Mass in
New Telecommunications Services

Until about a year ago the prospects for videotex seemed theoretical at best. Numerous trials, around the world, had produced what might be termed an "experimental" response from the user - in each case the trial service penetrated only to a nominal percentage of the target population. Then the French videotex effort Minitel, in the language of those who monitor new services, "took off." With terminals in place, today, for roughly 10% of all French telephone subscribers, with usage running at 37 million hours annually and with some 4,400 services vying for the users' favor, Minitel has become a dramatic contrast with the prior decade's unyielding effort to implant videotex among an entire society.

How does the interested observer - whether in a company, perhaps marketing or investment; or in government, a regulatory body or the courts; or in research, a laboratory or a university - how do we understand this phenomenon? Of course videotex - though Minitel has proven unmistakably remarkable - is only one instance that may provoke this query.

The same phenomenon may characterize other new telecommunications services. These include the wholesale reconfiguration of a national network, such as ISDN, or new forms of the local network, such as LAN's. For any one of these new kinds of networks how do we identify which, from among several trial efforts, will take off? What distinguishes the successful new network from the ones that sputter and fail to achieve lift?

Much of our recent understanding about the demand for networks has been based on experience with relatively mature telecommunications systems. However, a startup network faces a special circumstance that differs in one key respect from the mature situation. Since the fundamental raison d'etre of a network is to connect people, a new network - by its nature - requires a group of subscribers, if it is to startup. This need, for the critical mass group, surely separates the startup from the case of the mature system.

A mature system has necessarily moved beyond that point in its development where a critical mass initially assembled. As a consequence, new subscribers to a mature network can join one after another, rather than as a group. That indicates why, I believe, we should not expect to find evidence of network externalities in data from a mature system.[1]

This essay will focus on the distinguishing nature of network startup. Here we will inquire whether network externalities and critical mass can help to understand why some new telecommunications services succeed, when others do not.

Not all new telecommunications services have the characteristic of a new network. For instance, services that only replace existing mature arrangements - these include so-called "cream skimmers" who operate via the benefit of some universal service pricing umbrella - require little in the way of new behavior by subscribers, nor do they entail so much new plant and equipment on the supply side. It seems clear that new telecommunications services may have more, or they may have less, the characteristics of a fully new network.

The analysis in this paper is most directly applicable when the new service involves significant departure from past practices on both the demand and the supply sides of the network - when, that is, network "newness" is fullblown and network externalities and critical mass play their largest role. Full motion videoconferencing is a case in point; so would be an interactive broadband service to the home.

Voicemail involves less change, and therefore the analysis here is less directly applicable. At the conclusion of the paper, nonetheless, we will consider how critical mass in networks may also relate to mature systems and, so by implication, to those new services that are not so large a departure from mature arrangements.

Before we begin with critical mass, one additional prefatory note is in order. Externalities/critical mass are, of course, only one factor which determines whether or not a new telecommunications service succeeds. The service must, as well, fill some requirements of human communication. Obviously, whether such human needs are met - needs perhaps for visual tools, for message storage, for mobile reception, to suggest some possibilities - constitutes other, vital factors in the success or failure of a new service.

What puts externalities and critical mass in an unusual role, I believe it will emerge from our ensuing discussion, is their capacity, at least within a limited range, to determine whether these other factors are "accepted" in the first place, by a group of prospective users. Critical mass, it devolves,

is a sort of portal through which the other features must pass, if those features are to make their impact. But that is getting ahead of the story.

What is critical mass? How might we describe it?

Some framework

Can we characterize the economics of critical mass?

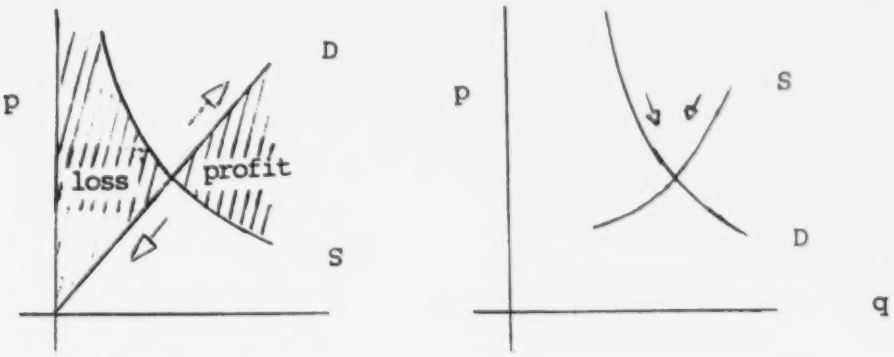
The existence of network externalities carries an essential message about the perception of value during a network startup: prospective subscribers perceive more value as the subscriber pool grows. Since the value of the network derives from its connectivity and since a larger pool of subscribers means more possible connections, value increases with size, at least to the limit of the community with which any subscriber wants to communicate. This suggests a demand curve such that, as the quantity of subscribers rises, so does price - that is, a demand curve which ascends with a positive slope, beginning perhaps at the origin.

(If, dear reader, you can at least momentarily suspend disbelief and resist to automatically reject this taboo thesis, its troubling character will, after some deliberation, become the central concern.)

In this special case of the network, the overriding function of price might be described as signalling, among the individual members of the subscriber pool, the fluctuating size of that group.

If we complete our picture with a supply curve - and I shall suppose increasing economies of scale* - we can define critical mass, at least in terms of its economics. We can also compare the case of network externalities with the more typical analysis in standard microeconomics.

Figure One



Network externalities Standard microeconomics

The "cross" between supply and demand, in the standard analysis, has - to produce the cross that emerges in the network case - rotated around its plane. What are the implications?

The standard analysis watches price mediate an adjustment process between supply and demand, one that progressively eliminates any discrepany - a surfeit or a shortage - that separates them. That promises convergence to a stable equilibrium at the anchorpoint of their intersection.

*Of course, the debate as to whether telecommunications is a declining cost industry is voluminous and lengthy. In particular, the rationale for the divestiture of AT&T held, as one fundamental, that the economies do not increase. Despite that, subsequent evidence has failed to settle the question in any conclusive way; in fact the increasing economies proposition, at least to some observers today, seems more than plausible. (And just to notice: any economies from declining costs are here compounded by a Chamberlinian sort of "local monopoly" that results from the network innovator's unique position.) Since the overarching topic in our main discussion is demand, this is not the place to pursue the debate about supply.

The network case however seeks a different style of analysis. With interdependence, price signals size among the demanders instead of mediating supply and demand. This, together with a monopoly supplier, switches the focus of the analysis to supplier incentives.

During startup of a network the entrepreneur faces losses since cost exceeds price - while in maturity the network provider reaps burgeoning profits when price increasingly spreads above cost. The result is disequilibrium forces as the incentives lead away from, not toward, the point of intersection.

In this disequilibrious arrangement, the point of intersection (which is an unstable equilibrium of just-matched revenues and outlays) defines the network's critical mass. Upon reaching that point the network no longer faces the losses endemic to startup. And if the network will then move off the unstable equilibrium, in the direction of a larger subscriber pool, the growing profit underwrites an internally generated "snowball effect," up the demand curve.

In addition to defining critical mass and illustrating its resistant-but-then-explosive character, this analysis also offers some detail on the startup campaign which any new service provider will face. The effort to overcome built-in inertia and reach critical mass will require the investment of a subsidy* in the amount of the shaded area on the left of the diagram.

*This phrasing, which purposely mixes the ideas of investment and subsidy, was chosen to cast startup investment in terms of the cross-subsidy debate.

Further, startup may begin with zero pricing; later, price can rise along with perceived value.*

This approach to the economics grew out of a suggestion that Charles Jonscher, in a discussion between us, made three years ago. About a decade before that, in a journal article, Jeff Rohlfs had derived the positive-sloped demand curve.[2] Charles Jonscher suggested the demand-supply cross, which I have dubbed the Jonscher cross.

So far in this approach, we have defined critical mass, conceptualized the scenario of passage from startup through to network maturity, and inspected some detail (of price and subsidy) that any provider of a new network will face in the contest against startup.

Clearly, however, there is a glaring inadequacy in the analysis, at least as it stands. Paul Samuelson has put it succinctly I believe. In his classic text he warns, in an injunction to the neophyte, that "the fallacy to be avoided like the plague is . . . : 'I have disproved the law of downward-sloping demand.'"[3]

Logic seems to argue - even persuasively to some, perhaps - that the

---

*Of course, the relevant pricing here is pricing for access. Since the central phenomenon at work is interdependence of choice - and since that turns upon the decision to choose access - pricing for access is the issue. When the primary role of access has been accounted in the analysis, then pricing for usage becomes a second order of consideration.

above account of network externalities can but make sense since demand is interdependent: positive, and growing, value depends on the positive, and accumulating, choices of other demanders. But opposed to this reasoning, apparently implacably, an increase of quantity with price strains beyond credibility, not to mention the decades of theorizing and empirical confirmation that now constitute demand theory.

Is the account above even sensible? Could the events of network startup, if inspected up close, reasonably be expected to transpire even loosely as outlined there?

That account hints at some underlying dynamic - some process whereby demand first must be "pulled," as it were, up its curve, then after reaching critical mass, pushes itself to maturity. Thomas Schelling has suggested an analytic device to inspect such events in their detail.[4] Ergo

More framework

The network externalities, again, carry a message: any individual prospective subscriber will actually decide to join only if some minimum number of the other prospects also decide to join. Each person has, as it were, a personal vision of what constitutes critical mass. That personal view may vary from individual to individual. Yet it seems, from the earlier pass through these events, that the network can succeed if it is able to gain just one critical size.

How do we understand the process of meshing the different individual choices of critical size - when that process appears able to produce coalescence for the group as a whole, at just one size?

In our description, here, each individual bases his/her decision on what he/she expects the others to decide. That means the effort to decide hinges on "watching the group," the other members in the community of actual/potential subscribers, to discern what the group choice may be. That is the case for each prospective subscriber. The outcome for the group turns upon, literally, everybody "watching, while being watched." Change, and finally critical mass for the group, float on shifting perceptions of what the group outcome may be.

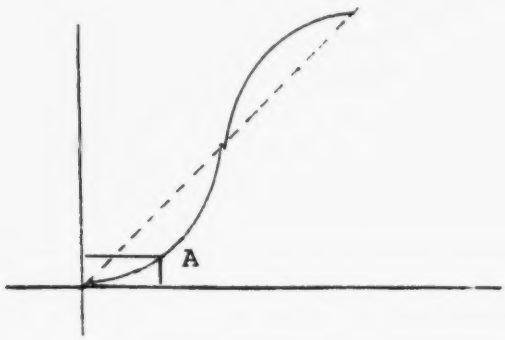
To portray this state of affairs we need to depict the interworking of separate, but interdependent, individual choices. To construct a picture, let us imagine that we learn from the prospective pool of subscribers what is the minimum critical size required by each individual. With that data we can plot a frequency distribution which will tell, for each minimum critical size, what part of the population would require just that size.

Now we convert those frequencies to a cumulative distribution. The cumulative diagram tells us, for any given critical size, how many prospects require that size or less, as a minimum - and therefore what part of the total population would actually want to join the new network, if it were expected to be that size. The vital question, of course, is whether the actual number who would join falls short of, or alternatively exceeds, the expected membership.

If the original frequency distribution is normal, we get this cumulative picture:

Subscribership - actual:
This number actually want to join

Figure Two



Subscribership - expected:
If this number are expected to join

The dashed 45° line - because it is a visual cue which tells us where the two variables equal each other - provides a handy reference by which to answer the vital question. In other words it helps tell at a glance, for any size expected membership: will actual results exceed (or just meet) and therefore serve to confirm the expectation of membership - or instead will expectations be dashed by the actual outcome?

For instance, a company's new intra-corporate videotex service could be at point "A" with membership expected to reach the point indicated on the horizontal scale, perhaps because the company has installed that many terminals. It is clear from the corresponding point on the vertical scale - or more simply, by noting that "A" is below the 45° line - that too few employees will actually use the service and expectations will not be met. Prognosis for the startup - given only these facts - would not be hopeful.

The operation of this process, though it begins with fixed individual needs for critical size, turns upon the repeated formation of a changing shared expectation about network size. (The details have been banished to an appendix.)

If we combine this conclusion with the earlier proposition about subsidy, we can see the purpose that subsidy should serve - without which, according to the earlier views anyway, the critical mass will not assemble. The subsidy must create a shared expectation that subscribership will be larger than critical mass. If that occurs, enough users will be attracted to trigger, simply by the fact of their presence, the growth yet needed for the network to reach maturity.

As the detail in the appendix confirms, this process moves in an incremental fashion through repeated disappointment, or satisfaction, of a shared expectation. Beyond such a step-by-step process, however, there is a larger phenomenon.

As the expectation changes incrementally, it progressively ratchets up or down. Eventually, though, these increments can cumulate and produce a more profound change in the group's expectation - in its outlook, you might say. The new network begins as an innovation and a curiosity - but if in time it does succeed, it is transformed and becomes a necessity, entrenched in the habits of the subscribing group.

To use commonsense terms: What begins as an attempt at a new network - say, videotex - may convert to being a new orthodoxy in communications. Whether this fledgling effort does actually emerge into orthodoxy depends upon whether the number of subscribers grows large enough to confirm, for the entire prospective subscriber pool, that the network has achieved legitimacy.

This shift from novelty to necessity, which is the larger-scale discontinuity in expectations, creates a new sense of value. It elevates an attempt at a new network to the status of essential commodity, sought after by the full pool of potential communicators, so that the once-new entrant finally emerges as the preferred mode of contact.

Perhaps it is the capacity to establish new economic value that is the real focus of interest in our larger investigation here. Certainly that is the noticeable result wrought by the lesser, incremental shifts in a shared expectation.

Price, as well, appears to serve this result. As a signal - within just the subscriber group - of changes in the size of that group, it can signal changes in value (tied, of course, to the changing size) which interdependent demanders are reflecting.

But surely, though, price is only one of numerous signals in this process of consensus formation. Consensus formation, it would seem, is the subject to understand. And the "signal" worth studying must be human communication - in its significant richness - the communication that comprises

an endless tug-of-war between potential consensus and potential diversity.* Nor will price likely be, I expect, the principal independent variable in that human exchange.

Recent work on standardization[5] addresseses some of these questions, though the game theoretic approach tends to emphasize the combative, over the consensual, aspects (despite efforts to ameliorate that) and also stylizes away vast realms of real human communication. In an earlier era, Robin Marris[6], and before him James Duesenberry[7], considered the range of questions here. Though Marris specifically rejected a positive-sloped demand curve as not a useful approach, he and Duesenberry both emphasized what I have termed value creation. At about the same time, J. R. Hicks touched on some of the preliminaries to these questions.[8] Finally, Schelling's model is one in a class of such models; several economists are pursuing a description of critical mass.[9]**

I have not yet, it seems certain, dealt adequately with the positive-sloped fly in the ointment. But before we press further, it seems prudent to check whether any of this may square with experience. The Minitel experience is one ready case.

---

*Communication, within the group, then becomes the primary focus. That one of our implicit topics here is technology for communications only spices the fray. My own inclination is to start with human events and then ask how the technology may interact in an evolution of the outcome.

**Since our main subject is demand, the question of critical mass in supply - a matter that is probably a full blown topic in its own right - will only be touched upon, and then only in a second brief appendix. A comment on diffusion, at footnote 10, also seems appropriate.

We have to ask, first, can we apply these ideas to videotex? Is videotex a new telecommunications service that involves externalities?

The Minitel experience: a case in point

The evidence from Minitel, as well as from a number of other videotex trials, is that person-to-person communication is a significant portion of the traffic. An example is the recent experience of five major services on Kiosk, the segment of Minitel which serves the general public. About half of the activity on these five main services has been devoted to various forms of "messaging," that is to person-to-person contact. Person-to-person entails, of course, the classic network externalities.

Perhaps the role of person-to-person on Minitel is most evident as it has emerged from communication that began initially as person-to-machine. Database queries (from person to machine, that is) on the subject of some niche interest - say, sailing - have made it possible, as a serendipitous side effect, to identify the subgroup of subscribers with a common interest in sailing.

Previously these individuals had primarily only print sources of information on sailing, such as magazines, sources which offered no easy means to identify others with a like interest. Once an electronic, person-to-machine "magazine" made it possible to identify that group, however, person-to-person began to flower.

If (albeit a bit brazenly) we put this into the terms that have been developed above, we get something like:

Person-to-machine serves to establish an expectation that a critical mass of like-minded discussants will "subscribe" to a related person-to-person sub-"network." More generally, the mechanism to implant this expectation involves interplay between separate critical mass scenarios for a pair of related videotex services, one person-to-machine the other person-to-person. Critical mass for the person-to-machine half of this pair, because there are not strong externalities on its demand side, has largely to do with its supply side.

. . . Minitel and pricing

Minitel is particularly striking because it has followed the contrary pricing path that the earlier analysis for externalities suggested. Service began with a zero pricing policy. As perceptions of value have risen with growth in the subscriber pool, it has been possible to begin charging prices which reflect the increase in value - in particular to begin charging for terminals, the surrogate for access to Minitel.* At present, about 8% of all terminals in place generate a fee.

. . . and subsidy/investment

Clearly, Minitel committed to a significant subsidy in its effort to

*As the foot at the bottom of page 7 noted, it is the price of access to Minitel, not the price of usage, which is relevant here.

surmount what would here be called its point of critical mass. From the American point of view there is a temptation to conclude that this subsidy has somehow been disproportionate and too large. In France such questions have stirred political controversy as well.* One point of comparison is the amount of investment that has been expended in the US, to establish videotex there.

Estimates of the expenditures on videotex in the US range from several hundred million dollars to as much as $1.5 billion - and of course these efforts have not yet borne fruit similar to the French experience. Since the French investment is probably not significantly larger, it is difficult to ascribe forbidding proportions to the French subsidy.

Free and blanket installation of terminals has proven the most remarkable aspect of the subsidy, in the eyes, anyway, of the world telecommunications community that has been watching. But it is worth noting that all the expenses of operation - terminals, but also all other activities - are surely important to an effective subsidy.

For instance I expect that communications, about the progressive spread of the network and about its widening usage, have played a key role for Minitel - at least that is what the ideas here would suggest. To the extent

*Again, the footnote regarding subsidy and investment at the bottom of page 6 is integral to the questions in this controversy.

that those "communications" were funded by marketing expenses of Minitel, they must be an essential element in the subsidy.

. . . and value creation

There is one point that was not drawn out explicitly in the discussion of value creation, but the implication is clear. Part of the challenge in critical mass scenarios is to change general perceptions of a new service, from a view that sees novelty to one that shifts and sees necessity instead. By using customer premises equipment that were as little different as possible from the not-at-all-novel voice telephone (though inevitably, different), also by giving the system a necessity service (the directory service),[11] Minitel reduced the level of this challenge to a minimum. That of course improved the odds for critical mass. In simpler terms, this might be described as "reducing fear of an unfamiliar activity."

The necessity directory service also played another important role, it appears. The presence of a necessity confers some guarantee of usage and, so, active subscribership. Thus incorporating a necessity component in the service design should help to cinch the expectation that there will be the largest possible network membership, hopefully a number greater than critical mass.

. . . value creation: forming the expectation

That is the core objective of course - to establish an expectation that subscribership will exceed critical mass. In fact, Minitel employed several means to implant the expectation.

. . . value creation: forming the expectation - sequenced introduction

One important means was a manageable campaign which introduced Minitel sequentially, in one geographical region after another. Critical mass could be begun in a region where prospects were strongest. As the campaign moved to progressively less fertile ground, early successes could serve as a growing foundation to support the later, more difficult efforts and thus help to build toward a regionally combined, nationwide expectation for critical mass.[12]

. . . forming the expectation - supply-side critical mass

Another fundamental means to the required expectation was, it seems likely, a separate critical mass on the supply side.* After initial conflict with the press, Minitel was able to bring together banks, newspapers and the computer industry, along with the PTT, into a combined effort that could orchestrate network, system, services and terminals.

Presumably, the function of this critical mass on the supply side would be to bolster the expectation that connectivity will be available. That expectation might best be confirmed when the prospective subscriber is able to try and to use a repertoire of new services. It seems likely that Minitel's

*Even though the primary discussion has not emphasized issues in supply, it seems likely that critical mass in supply, and its interaction with the parallel phenomenon in demand, may be useful for a satisfactory conceptualization of these events. (The second appendix introduces this idea, but only in summary form.)

coordinated suppliers could certainly provide a fair modicum of variety in new services.

How to capture the interaction between the two supposed critical mass phenomena? One possible approach: For Minitel the expectation that supply would be available "led" developments on the demand side. Indirectly though, that lead undergirded the expectation that subscribership demand could exceed critical mass, and so there was impact on both sides.

So Minitel lends itself to an interpretation along the lines that have, so far, been developed here. Perhaps you may even find that interpretation suited to a setting where it is easy to imagine a critical mass at stake - where a PTT takes decisions about a network that potentially can connect an entire society. But what about competitive situations, such as the liberalized and actively competitive environment in the US? How can critical mass assemble when the action is in pieces, instead of as a whole?

Critical mass for new telecoms services in competitive markets?

The essential issue, I believe, is how big will be what we might call the critical mass group. The competitive model is already apparent, I suggest, in Minitel's sequenced introduction, region by region. Sequencing means that the larger outcome is disaggregated into smaller, individual critical masses - which is, of course, the situation with competition.

So, then, how is the competitive case different from the single decision by a PTT? A major difference, it appears, is the separate critical mass on the supply side. In the PTT case, the expectation of available supply could support spread of the main target expectation on the demand side - spread, that is, from one group in a sequenced introduction to the next group.

The competitive situation, without such a supply-side link for its groups on the demand side, might fall back instead on excess capacity - supply, that is, which has been left over from some prior round of critical mass. In the competitive environment excess capacity from a prior round, as a surrogate for linkage from the supply side, may support a next round of critical mass, for some next group.

But how does critical mass assemble in a competitive situation? How does it get "over the hump," to form the shared expectation and create a new sense of value. Is some planner needed - to plan, say, the subsidy?

Though an intra-company planner, or manager, may decide on an investment that is intended to build confidence in a new service the company is offering, the spread inter-company, among previously- (or currently-) competing service providers, to interconnect different companies' different offerings - in the competitive situation - is surely a good bit more serendipitous than planned.

How about speed of that spread? How does the competitive situation compare with the PTT in reaching whatever will be universality for a given service? Perhaps the PTT with its planned environment can move the situation more quickly to a sense of new value - particularly if it guesses right as to

service and configuration and if, also, it does not lose a first mover's advantage to some second mover who learns from the PTT's experience.[13] Should we suggest, then, that planning may have some natural advantage?

I believe that would be premature. Instead, I think, I must return to the troubling proposition at the core of this analysis. Does it make any sense to suggest that demand can be upward-sloping?

The process of creating a sense of new value might be decomposed, for purposes of the analysis, by considering the constituent sub-groups which together comprise any whole subscriber pool. For the videotex case these are the sailing aficianados (to return to Minitel as an example), the chat-line users, the business subscribers, the professionals-at-home, and so on - each in his/her sub-group.

Yet more framework

At the beginning of network startup, one or a few of these sub-groups try out the new network. Each such initial sub-group must then, it would seem, traverse its own process of creating a new sense for what is the preferred network - must reach, in other words, its own critical mass, within its own membership, regarding this network innovation. Later, there could arise the prospect for critical mass across several of these sub-groups and, so, for communication across sub-group lines. Later even, if the new network has the potential to span a society, there might be the prospect for joining across these intermediate groupings and so to reach a universal critical mass.

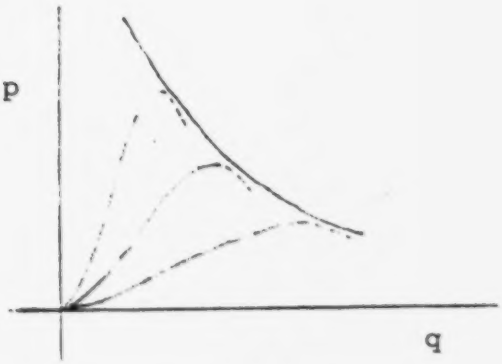
So far this is no more than a slighty formalized re-statement of sequenced introduction, the notion, it is worth recalling, that related the PTT to the competitive scenario. Does it serve the present question about the tilt of demand curves?

Let's look at our upward-sloping demand curve in terms of polar coordinates. And let's compare the three such curves that would represent the three levels of critical mass identified just above - sub-group, intermediate and universal. And finally let's suppose that the value added by that marginal subscriber who just edges the group up to the critical mass point is, when compared among these three levels, least for the universal case. Once, in other words, a telephone system has reached the verge of universal service for a society, the value of critical mass is less than the value of critical mass for the smaller, more compact intermediate stage; and that, in turn, is less than for the individual sub-group. This result, at first paradoxical, seems likely since the desire to communicate will be stronger within the individual sub-group than the desire to communicate with, say, anyone in the society.

In polar coordinates, the upward-sloping curve for the individual sub-group has the largest angle: the highest value for the smallest network size. The effect of "adding together" the sub-groups to cumulate to the intermediate stage, then "adding together" intermediates for universality, is, in polar terms, progressively to diminish the size of the angle. Finally - and finally to the point - when these three curves are drawn from the same origin, their separate critical mass inflection points perhaps begin to trace

out what is suspiciously similar to the downward-sloping demand curve.

Figure Three



Long Run Average Demand Curve

My name for the downward-sloping portion, as it accretes through the combining of progressively more sub-groups, reflects of course its conceptual symmetry with the long run average cost curve. Jeff Rohlfs presaged this formulation, in which the dynamic growth phase of a network innovation folds over into demand for a maturing network, when he posited that upward-sloping demand would bend and be followed by a conventional segment.[2]

One note: Of course the argument that the marginal subscriber adds less value for the universal case than for, say, the individual sub-group does not - by itself - indicate that the total value for critical mass at universality is less than total value for critical mass within the individual sub-group. In fact a lower value for universality, which after all is composed of sub-groups, may at first seem paradoxical, as already said.

But, if the same marginal analysis holds for each instance of the "adding together" of sub-groups - if "adding" one sub-group to another results

in a lower value being attached to communication across the two - then the seeming paradox dissolves. The marginal result for each accretion of a new sub-group leads to a steady progression down the usual - "long run average" - demand curve. And upon reflection, the conclusion that critical mass for universality holds less value than critical mass for, say, a small sub-group may begin to fit intuition.

And the Minitel evidence, to return there briefly for a last time? As already discussed, the price of access - which at present in Minitel is the fees paid for some of the terminals - has become positive with growth in the subscriber pool. Now we should complete that story: DGT's plan is that, with time, this charge will eventually diminish back again to zero. This pattern, of a transient rise in price during startup, followed by reversal and a decline in price with the onset of maturity, is of course the path described here.

But how does this square up with the usual view of the world? You likely say (if you are charitable): If this does have any usefulness, the upward-sloping portion is not a demand curve; it is the series of equilibria traced by a shift that produces a set of demand curves. Can we inspect that response?

Where is our conceptual tool for understanding the emergence of demand for a new, interdependently demanded service - where is a demand curve for the creation of new value in a network? Our current machinery, if we follow just above, would use the equilibrium points from a series of demand curves. Will Occam's razor be tempted to a bit of pruning?

It seems likely that our mental apparatus, in its efforts to conceptualize, is indeed less comfortable without static points, such as equilibria, to fix upon. But our understanding of demand did advance when demand was conceived as a shifting curve.[14] Can we revise the idea again, now to describe the dynamics of value creation in a parsimonious way? Can we develop a new conceptual tool - whether we call it demand curve or otherwise - that shows with simplicity the change, across that transient period, when interdependent demand emerges?

To the extent that may be possible, upward-sloping and downward-sloping demand may perchance have been got to lie down together with each other, at least here. What would such a formulation portend?

Since the industrialized economies have begun to liberalize their telecommunications sectors, a persistent question re-occurs: how to locate a boundary that will separate telecommunications activities that are properly market-based from those that are government-regulated.

Liberalization and the regulatory domain

Without much question, the US situation epitomizes this quandary. The FCC has now seen fit to convene, over the last 20 years, three major inquiries into the question. Each of these three episodes has proposed a new definition for the boundary line. The first two sets of definitions were each superseded on grounds that they had each proved to be inadequate, and the most recent

Computer Inquiry III proposals seem to offer little prospect for being more enduring or stable.

And this turbulence in the quest for a boundary line has proceeded quite separately from divestiture, the major act of liberalization which was itself a related boundary question. The divide between the regulated and the market-based seems clearly to obtrude as a basic issue. (Of course if we broaden our horizon, we know that the search for a regulatory divide sweeps a range of sectors, telecommunications one among them.)

The new view here - now with a focus on the demand side of events - could suggest a regulatory divide at the point of critical mass, between the growth phase in the first half and the period of maturation in the second part. The first half is, of course, along the upward-sloping segments and the second part is along the downward-sloping. Pricing in the first half would be subsidy-based (á la cross-subsidies from long distance to support the universal spread of voice telephony). Pricing during maturation, along the standard demand curve, could be conventionally competitive.

More basically, such a divide is dynamic - in its nature, it must move over time to accommodate rises and falls in new technologies, in new demand. That is to say, there is a reason why the search for an "equilibrated" regulatory divide is repeatedly frustrated. Does not that suggest the focus of attention needs to shift away from an equilibrated solution and onto the flexibility that would be necessary to support a continuing kind of change?

If this describes the divide between the regulated and the competitive, then it connotes a view of "regulation" substantially at variance with today's focus on quarantining monopoly from competition - which current view is of course the product of a supply-side perspective on the matter. The alternative, demand-side view of regulation would be concerned with the stage of development of the network.

Can we imagine, in a flight of fancy, what form this might take?

Just to conjure for a moment: Government is not a watchdog to industry; instead government and industry are co-facilitators who expect technology, and its demand, to continue to churn. Schumpeter must, surely, be the doyen of this view. As networks flourish, and later wither, the focus is on appropriate matching of subsidy and competition, respectively, to growth and maturity. Issues of "picking winners" do arise.

While price is one mechanism available, it is not the only one - and when used in a subsidy role, price does not serve the usual competitive/ allocative function. Negotiations, among all the actors, is another vital factor. So is competition among organizations, as an on-going spur, even when not on a price basis.

The position of the economically powerful, who might prey upon the less powerful, is a central issue - it is <u>the</u> central issue, in fact. Control is exercised however, not by rendering that power impotent, but by harnessing it.

To be sure, this is a fantasy in today's American economy. Perhaps a few other capitalist economies are be a bit closer to such a conception.[15]

The issue of where, and ultimately how, to regulate is one question to which the demand-side conception might address itself. There is another, intimately related problem which liberalization has raised. As networks have moved toward maturity and toward competitive, cost-based pricing, there is an increasing conflict with pricing for universal service.

With a migration to cost-based pricing, higher prices to those subscribers who cause higher costs comes into increasing conflict with an objective, held by some in the policy dialog, that all subscribers should be maintained on the network. Of course, one of the basic observations of our analysis is that the "snowball" effect, during growth to maturity in the second half, has built into it a natural tendency for the network to remain at its "universality" equilibrium. However price differentials, assessed against higher cost subscribers, does obviously countervail this tendency.

Liberalization and universal service pricing

Put most simply, the demand-side approach raises the possibility for treating the critical mass group as the unit of purchase in the transaction between providers of networks and users of telecommunications services. The relevance of this to universal service pricing works itself out, fairly obviously, as follows.

First, to begin with the contrasting case: If the largest critical mass group is small relative to the society - if, in effect, universality for the entire society is not really meaningful or important - then the distinctions drawn in this paper lack force. Group consensus formation does not have significant impact on the economic outcomes since, given only small groups, the network analysis approximates the standard microeconomic analysis.[16] And without universality for the society, there is no conflict between competitive pricing and universality. Pricing can be determined in the usual competitive way: with differentiation that recognizes variety in communications services (and related costs), as demanded by many small groups.

On the other hand: When the critical mass group is large and universality a prospect, there is potential for the competitive pricing versus universality conflict. If, however, network services are purchased in units that are the size required to accommodate the critical mass group, then the conflict may dissipate. Pricing this critical mass-sized unit - the pricing per subscriber, that is - can, with this arrangement, follow typical universal service practice and average by subscriber.

To make such a proposal is to do no more than recognize that, in demand for the network, there can be a counterpart to natural monopoly in supply. Or to put it a bit differently, this recognizes a natural monopsony counter-balanced to a natural monopoly.*

*And thus, with this new view, we see why the monopolist does not control price (as in the standard analysis). Instead, there is a monopsony/ monopoly seesaw. So then we see, further, that it is appropriate to speak of a supply curve, even in this situation.

On a sober note, this proposal does fail to deliver the promise of real competition during the downward-sloping maturity phase. Indivisibility in demand only revisits all the problems of indivisibility in supply and natural monopoly, certainly including the way large-scale "lumpiness" precludes any real prospect for competition.

On the brighter side however, we are now in a position to look more squarely at the signalling function of price in the case of the new network. Now we can look alternatively at first the rise, then the maturity of a network. Here we hear, I believe, a refrain on the "paradox of value" which traces back at least to Adam Smith.

While price signalling during startup may really be just within the group comprising potential demand (and, at that, not the most important signal in a consensus formation/fragmentation process), price during the maturity phase nonetheless returns to its more usual function of mediating between supply and demand.

But if the balance in maturity is really between two sides each with natural scale economies, does price signalling in maturity come down to the same marginal result which made it possible for neoclassical microeconomics to resolve what came to be known as Smith's paradox of value? It would appear, instead, that price in maturity serves to reflect relative balance between the two sides. And, intriguingly, the role that price plays during startup means that, during that time, it takes on what is often thought of as the common-sense meaning of price, namely value to the consumer.

An underlying question; organization to support this process

Earlier we asked whether a planned approach might have some advantage over competition, given the state of affairs described here. But subsequently, the flexibility needed to support change has turned out to be important. So just which is it - market or hierarchy? I think probably neither - or more to the point, both, for this network case. At least, that is what these competing requirements would suggest. To put it another way, the mixed economy may be the first best, not a second best, solution.[17]

One possible such intermediate case, one that nicely combined elements from both opposites, may be the program in Japan which created universal service in a few decades. And perhaps Japan's financing of the facilities for that universal service illustrates, at least in one small way, the intermediacy of the approach. While there was significant coordination and funding from the centralized authority, individual subscribers in the market also participated, through their purchases of bonds. In fact such financing provided close to half the investment required.

And just what is this saying?

The prevailing general view of economic activity, now so carefully worked out, has surely provided an elegant understanding of an internally consistent and self-generating resource allocation. And what of a process of value creation? Could that be a demand counterpart to the resource side, but one that injects an inherent dynamic into the interplay between the two - that

underwrites a continually changing ebb and flow of technology-with-demand, a repeated undulation in value and the wherewithal to provide that value?

If we are to understand and enjoy, alongside our neighbors, the full fruits of evolving new technology/new demand, certainly we need conceptualization thereto.

And a rising demand curve (perhaps not through the origin and of shape unknown) dear reader - where are your sympathies now? I cannot expect you are convinced. If you look around you though, I suggest you may find even common cases that are not adequately explained by a shift in curves, particularly in new technology situations - at least that is what has happened to me - though I hesitate to admit that these are not always instances of a pure network externality.

Appendix A:

Network demand adjustment, in detail

If we look up close at Figure Two:

- In the region below the inflection point of the curve, it is clear that too few of the population actually respond to match any given level of expected membership. [This outcome depends of course upon the data, i.e., the normal distribution, with which we began originally.] So those who do respond find the actual network sizes are inadequate to sustain their own, individual requirement as to critical mass. Any fledgling network that operates at sizes in this region will disappoint those hardy souls who do give it a try, and the effort will slide back to oblivion (lacking, that is, exogenous tweaks).

- Above the inflection point, on the other hand, the network size always exceeds expectations, even for prospective subscribers with the next higher required critical mass - and so these next subscribers also find reason to join. That creates the self-powered snowball which generates growth to universality.

- The inflection point itself is, of course, the point of the group's critical mass, which any startup effort must surmount. That point of critical mass is in equilibrium, but it is unstable: any little tip to one side or the other - the loss of a few subscribers for extraneous reasons, or the gain of a few - will unbalance the situation and the process heads off, pell mell, to the destiny toward which it was tipped. Zero subscribers and universality are both also points that are in equilibrium; and these equilibria, in contrast with the critical mass, are stable.

As I suggested in the text, changes in the prospects for a new network - whether that means a slide back to the starting point or a climb to universality - might be described as each prospective subscriber "watching, while being watched." The details of that can now be specified.

To begin, we inspect one cycle of change in the pool of subscribers:

- Say we inspect some single slice, executed at any point along the history of our new network. (Since our focus in this closeup inspection is on change, the only "slices" that we do not consider here are the three equilibria.) We find that the group holds an expectation about what the network size will be, with some people having selected membership, while others still hold back.

- There are a subgroup for whom the shared expectation is particularly salient - those whose personal requirements for critical size are near,

or equal, the present level of expected membership. These individuals must confirm whether their needs have been, or might be, met by the existing state of affairs. They do so by comparing actual network subscribership against the critical size which they personally require. And because personal critical sizes, for this subgroup, approximate the current group expectation - during this particular cycle of change - their check serves as a check of the group expectation against actual.

- o Changes, individually and for the group, work out as follows: When the network is below critical mass, the key individuals are the current subscribers who find that their critical sizes - and thus, by approximation, the group expectation - are not met. Alternatively above critical mass, the key individuals are the prospective subscribers who find that their critical sizes, again along with the group expectation, are exceeded. These individuals then desert, or alternatively join, the network. To reflect this change in fortunes, the group expectation must adjust and it ratchets down, or ratchets up, to accord with this most recent shift in the situation.

- o Then, with a new expectation established, key individuals - who are different from the key people in the round just completed - check once again and . . .

- o For instance: If the expectation is 35, but the actual is only 30, then present subscribers with critical sizes between 30 and 35 desert the network. The goup expectation then adjusts to reflect the new reality of 30. But [at least with the curve from our original, normally-distributed individual needs for critical size] after those desertions there will, again, not be enough subscribers to support the revised expectation, so . . .

Change for the network, its ebb and flow, turns on a basic interplay between two factors, one constant the other in flux. On the one hand, the distribution of different individual needs for critical size, within the group, is a fixed element. On the other hand, the group's shared expectation about network size continues to change (and, as with any expectation, tends to lag the situation it mirrors by at least a period). Outcomes for the network - the seemingly inexorable climb, or demise - grow from the interaction between the individual critical sizes and the changing group expectation.

The speed of the adjustment process - in other words, the speed of change - depends on whether perceptions of the changing actuals are fine- or coarsely-grained and on how well "lubricated" are communications within the group. Those communications underpin the test of expectations - to match the perceptions of actuals against expectations. The communications also underpin what is the decisive step, namely the shift in expectations - here the communications must match perceptions of the <u>change</u> in actuals against the current, but soon-to-be-obsolete, expectation.

This approach emphasizes the importance of communications, within the group, for the speed of the adjustment process. The central character of

those communications, the "watching, while being watched," means that in a real sense outcomes "float" on the perceptions of, and as a creation of, the group. Since styles - and efficacy - of communications vary significantly by group (or by culture), this floating creation of the group may move quickly, or may move slowly, to establish new economic value, the effect of the larger shift in expectations. Some groups may evolve new values, where others are not so able.

Appendix B:

Critical mass in supply

The critical mass phenomenon appears to apply to demand separate from supply. As we described passage through a critical mass using the dynamics of expectation change - a description which otherwise seemed simply to elucidate the earlier supply-and-demand picture - we found no reason to refer to supply. The critical mass process worked in demand by itself, without the need to invoke events from the supply side.*

The Minitel experience also suggests a separate set of, in effect, externalities in the development of supply - in the development of the technology - especially where physical networks, such as for telecommunications, are at stake. It is not a long step from there to imagine that the separate critical mass scenarios in demand and supply interplay with each other. First one side, then the other would take the lead and create a change in expectations, which would then impact both sides.

*Except of course for implicit reference to the existence of supply, not a small issue to which the text subsequently returns.

Footnotes

1 The economist's notion of externalities can be construed, I believe, to denote the process whereby a group, rather than an individual, makes (or forebears to make) economic choices. Construed in this way, the new denotation derives - albeit awkwardly - from the original framework which puts the individual decisionmaker at the center of the conception of externalities. With the individual as starting point, choices by a group are "external" to any individual. Of course, it is not necessary to espouse centrally planned economies or for that matter oppose a paradigm of individual choice on intellectual grounds, to accept that groups, composed of individuals, do make choices - as occurs in the case of network externalities.

2 Rohlfs [1974].

3 p. 59, 11th ed.

4 Schelling [1978], Chapter 3.

5 Selected examples include Farrell and Saloner [1986], Katz and Shapiro [1985], Berg [1984]. Besen and Johnson [1986] offer a fairly complete review of this work, including several I have not listed.

6 Marris [1964], particularly chapter 4.

7 Duesenberry [1949], particularly chapter 6.

8 Hicks [1956]. See particularly sections 3 and 4 in his summary. Of course, a more historical provenance must reach at least back to Pigou.

9 Just one example is Brian Arthur.

10 Suffusing the desire, throughout a community, for some new technology (which is another way to describe value creation) is of course the diffusion of an innovation. If the horizontal scale in Figure Two is renamed "time" - which would be an appropriate alternative - the critical mass process nicely reproduces the S-curve that typically is found in research on diffusion. That seems at least suggestive of the role which externalities (perhaps in a weaker form but nonetheless there) may play in the spread of many sorts of innovations, even those where interdependence appears to be absent.

But this is not the place to inquire whether the analysis here may apply to investment/subsidy in any generic new product. Instead I will just note that Bass models, and their progeny, in the diffusion literature, tend to keep separated - with their distinct innovators and imitators - what here is treated more as a gradient

11 Of course the full history is that the paper directory was eventually reinstated for Minitel customers. But it is easy to suspect that during the time Minitel was still the sole directory, in the early stages, the necessity aspect made itself felt when that counted most - when critical mass was most at stake.

12 With this sequential growth pattern, the ability of the supply side to be appropriately responsive turned upon there being a flexible network architecture. To create the decentralization essential for this flexibility, Minitel pioneered some of the innovations necessary for decentralized operations on such a scale.

13 Of the work on this endlessly debated question, Farrell and Saloner [1987] have a recent intriguing piece that illustrates how planning may have an advantage in speed.

14 A summary of this history by Elster and Hylland [1986], page 1, is succinct but pungent.

15 On another tangential note: At a metaphorical level if the last diagram is seen as growth demand emerging into mature demand and is imagined on a grand scale - at the level of the society - then liberalization itself is represented by the process of sectors emerging from the shelter of more centralized subsidy/support and into the arena of disposition through competition.

16 So small groups approximate the standard competitive situation. That, however, should not be confused with substituting the critical mass group (large or small) to be the "q" in Figure One, in place of the individual subscriber. To make the critical mass group the basis for counting quantity serves again, I believe, to illustrate how critical mass formation may "fold over" into mature growth. If whole groups constitute individual "q", the groups proceed down the usual demand curve ("adding together", of course, to be larger groups at higher "q"). On the other side, if activities internal to a group are inspected up close (and "q" again becomes a count of individual subscribers), the transient, upward-sloping critical mass process emerges to detail any single point on the downward-sloping curve.

17 Again, see Farrell and Saloner [1987] for some parallel notions that the "hybrid" case works best.

References

Berg, S. V. "Technological Externalities and a Theory of Technical Compatibility Standards," mimeo, University of Florida, revised November 7, 1984.

Besen, Stanley M. and Leland L. Johnson. Compatibility Standards, Competition, and Innovation in the Broadcasting Industry. Santa Monica: Rand, November 1986.

Duesenberry, James S. Income, Saving and the Theory of Consumer Behavior. Cambridge, Mass.: Harvard, 1949.

Elster, Jon and Aanund Hylland. Foundations of Social Choice Theory. Cambridge, England: Cambridge University, 1986.

Farrell, Joseph and Garth Saloner. "Coordination Through Committees and Markets," Working Papers in Economics E-87-25, Hoover Institution, Stanford University, June 1987.

Farrell, Joseph and Garth Saloner. "Installed Base and Compatibility: Innovation, Product Preannouncements and Predation," Working Paper Number 411, Department of Economics, MIT, revised February 1986.

Hicks, J. R. A Revision of Demand Theory. Oxford, England: Oxford University, 1956.

Katz, Michael L. and Carl Shapiro. "Network Externalities, Competition, and Compatibility," American Economic Review, vol. 75, pp. 424-440, 1985.

Marris, Robin. The Economic Theory of 'Managerial' Capitalism. New York: Free Press of Glencoe, 1964.

Rohlfs, Jeffrey. "A Theory of Interdependent Demand for a Communications Service." The Bell Journal of Economics and Management Science, vol. 5, pp. 16-37, Spring 1974.

Schelling, Thomas C. Micromotives and Macrobehavior. New York: Norton, 1978.

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

REGULATION, DEREGULATION & COMPETITION
Chair: Jeffrey Rohlfs

Neal C. Stolleman: A Comparison of Rate Base Rate-Of-Return and Ceiling Price Regulation

The FCC has proposed replacing rate-of-return regulation for dominant carriers with a system of price ceilings for interstate services. It appears that the system would be implemented first for AT&T, with application to exchange a carrier operations coming at a later time. This paper represents a comparative static analysis of these two alternative regulatory systems from the perspective of a multioutput, profit maximizing exchange carrier. Evaluations are made in terms of the effects of these systems on composition of outputs and inputs, and the distribution of innovational effort.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Edward Zajec

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

013180

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

A COMPARISION OF RATE BASE RATE OF RETURN

AND CEILING PRICE REGULATION

NEAL C. STOLLEMAN

GTE SERVICE CORPORATION

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Edward Zajec

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Neal C.Stolleman

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

IR013180

I) INTRODUCTION

This paper represents a comparative static analysis of a multioutput, profit maximizing exchange carrier operating under two alternative forms of regulation: 1) rate base rate of return, and 2) ceiling price regulation. Analytical methods are used to evaluate and compare the distortions these regulatory forms create in terms of output/input composition and distribution of innovational effort.

A number of qualitative results are obtained. Regarding rate base rate of return regulation, the analytical model predicts the already well known result that the firm will employ excessive amounts of the resources included in the rate base. In addition, the output mix becomes distorted when there are jurisdictional differences in authorized rates of return and the cost allocation process is sensitive to the composition of output. And, depending on whether R&D expenditures are in the rate base, the distribution of scarce innovational effort between network and non-network operations can be skewed. In particular, excessive reliance may be placed on using more nominal resources as a means of raising effective network capacity, instead of innovational effort to enhance the productivity

of the nominal resource base. And, just the reverse may be true with respect to non-network operations.

Regarding ceiling pricing regulation, the firm's output mix is distorted by the need to satisfy the constraint of an artificially low price on one of its product lines ("core service"). An apparently counter intuitive result is that, under this regulatory system, the output mix becomes more intensive in goods that are substitutes with the core service and less intensive in goods that are compliments with it. The intuitive reason for this anomaly is that more substitutes and less compliments are consistent with keeping the price of the core service artificially low. And, if the price cap is lowered when innovation raises the productivity of resources utilized by the core service (as has been suggested), the level of innovation in that particular network segment is retarded.

In order to go beyond qualitative analysis requires building new or modifying existing simulation models to incorporate the model structure discussed in this paper.

From the perspective of a profit maximizing firm concerned with obtaining high returns for its shareholders, the preferred regulatory system is the one that yields the highest profit level. This is an empirical issue. Moreover, as new technological developments erode the traditional monopoly of the local franchise, the markets for exchange carrier network services become more competitive and contestable. Hence, the goals of profit

maximization and economically efficient use of resources tend to merge into one another.

II) MODEL STRUCTURE AND ASSUMPTIONS

Outputs and Inputs of the firm

The exchange carrier is represented as a multioutput firm. Its product set consists of the following categories:

X1 = single line subscriber access (into switched network)

X2 = multiline subscriber access (into switched network)

X3 = local usage

X4 = interoffice usage (toll or switched access)

X5 = private lines

Single line subscriber access refers to residential or business customers who rent one local loop (RIs and BIs). Multiline subscriber access refers to business customers who connect to the network using PBX trunks. This category would also include shared tenant services which concentrate the traffic of smaller residential or business customers onto PBX trunks. Local usage is that traffic which is switched just within a local exchange area. Interoffice usage is that traffic which has to be switched out of the local exchange area and transported to an interexchange carrier's point of presence, or to an exchange carrier's intraLATA toll network.

Private lines are dedicated facilities that connect customers to interexchange carrier or exchange carrier toll networks . These connections do not require the switching function to be performed. Private line traffic may flow into an interoffice switched network, as is the case with service bypass. Alternatively, the traffic may flow into long distance private line networks. In either case local usage is not carried on these circuits (in reality, "leaky" PBXs do permit terminating traffic on dedicated circuits to filter into the local network). With the advent of Software Defined Network, the same private line trunk group connecting the customer to the interexchange network can carry both traffic destined for long distance private line transmission as well as traffic destined for the long distance switched network.

In order to produce its output set, the exchange carrier must utilize the services of different kinds of telecommunications network capital as well as non-network resources. The input set of the exchange carrier consists of:

Z1 = Line side network access capacity (feeder,distribution plant)

Z2 = Switching capacity (local and tandem)

Z3 = Trunkside plant (transport facilities)

Z4 = Non-network resources (Overhead, operating expenses)

Line side network access capacity refers to the physical amount of economic capital needed to support the quantity of single or

multiline access and private line connections. Although traffic streams carried by RIs and BIs may be segregated from traffic carried by high capacity special access or private line circuits, physically, all of these circuits may share the same sheathing or conduit infrastructure.

Switching capacity is also shareable between local and interoffice usage, and associated costs depend on aggregate system traffic. Transport facilities linking central offices or access tandems with toll switches support both switched and dedicated services, although the product (and often, customer) specific traffic streams would be segregated within the same physical facility.

The final input category, non-network resources, is defined as a composite of administrative overhead, variable expenses, and so on, which, from the viewpoint of the exchange carrier's economic system, represents the production of a quasi-public good or social infrastructure which supports all product lines.

The following table summarizes the relationship between the outputs and inputs of the firm:

OUTPUTS	INPUTS			
	Z1	Z2	Z3	Z4
X1	Z11			Z41
X2	Z12			Z42
X3		Z23		Z43
X4		Z24	Z34	Z44
X5	Z15		Z35	Z45

Each of the Zi's represents the maximum capacity of each type of resource. The table summarizes how utilization of the capacity is distributed across product lines. For inputs 2, 3 and 4, maximum capacity is a flow rate per unit time. For input 1, maximum capacity is a stock variable.

The table, and the preceding discussion, indicate that the maximum capacity of each input is divisible in use among different product lines, and, for this analysis, all available capacity is assumed utilized. These resource constraints can be expressed as:

---

1) $Z1 = Z11 + Z12 + Z15$

$Z2 = Z23 + Z24$

$Z3 = Z34 + Z35$

$Z4 = Z41 + Z42 + Z43 + Z44 + Z45$

Capacity costs

The cost to the exchange carrier of producing a given vector of network resource capacities is expressed as:

2) $C = C(Z1,Z2,Z3)$

There are several points related to this formulation. First, "C" represents the cumulative, capitalized costs of constructing the existing amount and type of available network functionality. It is therefore a stock variable, and is analogous to the rate base to which an authorized rate of return is applied under the current regulatory system. The purpose of writing this (stock) cost as a function of the vector of network functional capacities is to reflect the cost characteristics of the construction process. That

is, the marginal stock cost of increasing maximum switching capacity is dC/dZ2, and depending on economies of scale in construction, this derivative may decline as Z2 is increased.

A related point is that the marginal costs of adding capacity are long run costs, because full utilization of network resources is assumed. For simplicity, it is also assumed that network capital is divisible, as opposed to being lumpy, so that it can be added continuously.

Furthermore, equation 1 reflects economies of scope which may arise in the joint production of several different kinds of network functionality. For example, the cost of increasing network access capacity includes both outside plant costs, as well as the cost of terminating the lines on the central office switch. Thus, the extent, or lack of, switching investment can affect the cost of increasing access to the network. The same is true of transport capacity, because outside plant must be combined with circuit or radio equipment which feeds into a switch.

Another characteristic of equation (1) is that costs depend on the aggregate amount of each type of network function, and are independent of which product lines utilize the capacity. For example, the cost of increasing switching capacity, dC/dZ2, is independent of whether the extra capacity is used for local or toll calling. (This is strictly true if reference is being made to a very condensed network which handles both kinds of traffic. In a

dispersed network with many local and access tandem switches, the distribution of local and toll traffic through the network would be different. Long run marginal capacity cost at a particular switch would be independent of traffic mix, but, the number of switches affected by an increase in just local traffic, for example, would depend on the traffic distribution throughout the network. Therefore, in an aggregate sense, marginal increases in switching capacity could depend on which kind of traffic was the driver. To effectively deal with the more atomized version of the network would require abandoning this generic analysis in favor of engineering simulation studies).

Once installed, the network capacity vector, (Z1,Z2,Z3), is available for use over time (neglecting depreciation). But given the comparative static nature of this analysis, the goal of the firm is to maximize the instantaneous rate of profit. Thus, the stock cost, "C", must be converted into a flow to reflect the instantaneous cost of consuming capital services. This is accomplished by multiplying "C" by the user cost of capital, r:

3) $$rC(Z1,Z2,Z3)$$

Non-network resources will be considered as variable inputs, with constant cost of acquisition P4. Total variable costs are therefore (P4)(Z4).

Technical Innovation

Total cumulative expenditures on research and development (R&D) will be defined as "M". This stock cost is a proxy for the state of technical knowledge of the firm, as well as a measure of total dollar investment in technical knowledge (depreciation of technical knowledge is neglected). As in the case of network capacity costs, the stock of cumulative R&D expenditures is multiplied by the user cost of capital, r, to convert it into a flow rate.

A given level of technical knowledge, M, can be distributed to different parts of the exchange carrier's operation for purposes of enhancing the productivity of its nominal resources. All of the accumulated knowledge represented by R&D expenditure, for example, can be entirely directed at enhancing line side network access, at increasing switching, or at a combination of the two, and so on. Intuitively, there is an expectation that devoting a larger and larger fraction of finite knowledge to improving productivity in just one sector eventually leads to diminishing marginal returns, with concommitant increasing marginal returns to innovational effort in those sectors in which the supply of such effort is reduced.

The feasible distribution of innovational effort open to the exchange carrier, for a given state of knowledge, M, is summarized by the following Innovation Transformation function:

4) $H(M,a1,a2,a3,a4) = 0$

The variables a1, a2, a3, a4 are called efficiency parameters. They correspond to each of the Zi inputs discussed earlier. Their purpose is to convert each of the Zi's, which are measures of nominal capacity, into measures of effective units of capacity. That is, the nominal amount of line side network access capacity is Z1. The effective amount of line side network access capacity is (Z1 a1). Thus, a1 measures the number of effective units of a resource per nominal unit. In this example, the size of a1 could depend on the number of pair gain systems installed.

Similarly, a2 measures the amount of effective switching capacity per nominal unit of switching capacity, so that (Z2 a2) represents total effective capacity. The size of a2 may depend on the diffusion of digital switches in the network, as well as software improvements within the digital technology. The transport efficiency parameter a3 would depend on fibre optic technology and improvements in the associated electronics. Finally, regarding the composite variable resource, a4 might reflect innovations in office equipment, hiring procedures, or internal training programs.

An important aspect of equation 4 is that it represents a constraint which limits the ex-ante choices open to the exchange carrier. If the stock of knowledge, M, remains constant, innovational effort aimed at line side network access, for example, cannot be increased unless there are simultaneous reductions in planned innovational activity elsewhere in the system. If M increases, however, productivity in all aspects of the firm's operation can be

improved.

The exact ex-ante distribution of scarce innovational effort which the firm chooses to maximize its profit, however, will depend on the demand and cost conditions facing the firm, as well as the type of regulatory system in place.

Feasible Production Plans

A feasible production plan is a combination of outputs and inputs the firm is physically able to acheive, given the capabilities and limitations of its production process, and the state of its technical knowledge. The exchange carrier produces several outputs, each of which may utilize portions of the (effective) capacity of several types of resources. A compact way to represent the feasible production plans is the transformation function:

5) $$F[\, X1, X2, X3, X4, X5, Z1a1, Z2a2, Z3a3, Z4a4 \,] = 0$$

This relationship summarizes the ways in which the various inputs, measured in effective units of capacity, can be transformed into different combinations of the outputs on the efficient frontier of the production surface. For example, for given amounts of all effective inputs and all outputs except, say, number of multiline access connections, the formulation solves for the maximum amount of multiline network access connections which can be accommodated. Alternatively, for given amounts of all outputs and all inputs,

except transport capacity, the solution of the equation indicates the minimum amount of transport that will sustain the existing output set.

The exact production plan chosen by the exchange carrier to maximize its profit will depend on demand and cost conditions, and the type of regulatory system in place.

III) PROFIT MAXIMIZATION IN THE ABSENCE OF REGULATION

Objective Function

The objective of the firm is to maximize the difference between its revenue inflow and its cost outfow, subject to its production and innovational constraints. This problem is expressed mathematically as:

6) $R(X) - rC(Z) - rM - (P4\ Z4) + L1\ F[X,Za] + L2\ H(M,a)$

In this expression, R(X) stands for the revenue inflows the exchange carrier can receive, which depends on the set of outputs produced (for ease of notation, X represents all of the possible outputs, instead of writing X1,X2,X3,X4,X5. Similar shorthand notation is used for the Zi's and ai's). The relationship between revenue and different possible output sets implicitly contains all the relevant information concerning own and cross price elasticities of demand among the outputs.

Costs are then subtracted from revenue. As was discussed above, costs include the cumulative expenditures on network capacity and R&D, both converted into flow rates. The last component of cost is the composite for all non-network (variable) resources.

First Order Conditions and Interpretation

The first order conditions for profit maximization are shown below. Satisfaction of these conditions determine the optimal outputs, X, quantities of nominal resources, Z, the level of R&D expenditure, M, and the size of the efficiency parameter for each resource, a.

First order condition for output:

7a) $MR(i) + L1\ F(Xi) = 0 \qquad i = 1 \text{ to } 5$

MR(i) is the marginal revenue of the iTH output. L1 is a lagrangean multiplier, and F(Xi) is the derivative of the transformation function with respect to the iTH output (conventionally defined as a positive number).

First order condition for nominal input (network capital):

7b) $-rC(j) + L1\ F(Zj)aj = 0 \qquad j = 1 \text{ to } 3$

Here, C(j) is the marginal stock cost of increasing the capacity of network capital Z(j). In turn, this cost is converted into an

annuity stream by the user cost of capital. Ll is the same lagrangean multiplier. F(Zj) (defined as a negative number) is interpreted as the derivative of the transformation function with respect to effective units of Zj (i.e., Zj aj). To obtain the derivative with respect to nominal units, F(Zj) must be weighted by aj, the number of effective units contained in each nominal unit.

Combining these first order conditions yields:

7c) $MR(i) + [\ rC(j)\ F(Xi)/F(Zj)aj\ ] = 0$

which is equivalent to:

7d) $MR(i) - [\ rC(j)/MP(Zji)aj\ ] = 0$

obtained by differentiating the transformation function. Further, MP(Zji) is the marginal product of an effective unit of resource Zj in the production of output Xi. This marginal productivity is weighted by aj to convert it into the marginal productivity of a nominal unit of resource j.

Thus, profit is maximized when the marginal revenue of each output Xi is set equal to its marginal cost, or, alternatively, when the marginal factor cost of each resource is equal to its marginal revenue product in all the sectors where it is utilized.

The interpretation in 7d is in terms of flow variables. If the

expression is divided through by r, then the interpretation is that the present value of marginal revenue streams must be equal to the marginal stock capacity cost incurred in order to accommodate the higher flow rate of output.

The first order condition for the level of R&D expenditure is:

7e) $-r + L2\ H(M) = 0$

where L2 is the lagrangean multiplier associated with the innovational constraint, and H(M) is the derivative of that constraint with respect to the level of R&D expenditure (defined as positive).

The first order condition for the optimum size of efficiency parameter aj is:

7f) $L1\ F(Zj)Zj + L2\ H(aj) = 0$

where, as before, F(Zj) is the derivative of the transformation function with respect to effective units of resource Zj. This derivative is then weighted by the amount of nominal Zj to reflect the change in the number of effective units associated with a one unit change in aj. To place an intuitive interpretation on these conditions note that from 7b, the lagrangean multiplier L1 is equal to rC(j)/F(Zj)aj, and from 7e, the multiplier L2 is equal to r/H(M). Substituting these terms into 7f yields:

7g) [rC(j) Zj / aj] + [rH(aj) / H(M)] = 0

and after differentiating the innovational constraint and rearranging terms the optimal amount of aj is determined by:

7h) rC(j)/aj - rM(aj)/Zj = 0

M(aj) is the marginal change in R&D expenditure needed to obtain an increase in the value of aj, holding all of the other efficiency variables constant.

The interpretation of 7h is as follows: Effective capacity of resource j can be increased either by producing more nominal units (Zj), at a given level of efficiency, by increasing the efficiency (aj) of a given number of nominal units, or by a combination of the two methods. At the margin, the cost of producing the last unit of effective capacity should be the same regardless of which method is used. If this condition were not satisfied, then total costs could be lowered by shifting the production of the last effective unit to the cheaper method. The first term in 7h is the marginal cost of a nominal unit of resource j, divided by the number of effective units contained in that nominal unit, which yields a marginal cost per effective unit. The second term is the marginal R&D cost of increasing the efficiency parameter for resource j, divided by the total nominal amount of the resource. Therefore, the marginal R&D cost is spread over the total increment in effective units resulting from increasing the value of aj. Optimality requires that both terms be equal.

The next section will investigate how these first order profit maximizing conditions are affected under two regulatory systems: rate base rate of return and ceiling price regulation.

IV) PROFIT MAXIMIZATION UNDER RATE BASE RATE OF RETURN REGULATION

Overview of Current Regulatory Framework

Under the current system of cost allocation and recovery rules, local loop and line termination costs are apportioned between the intrastate and interstate jurisdictions on the basis of a percentage factor that essentially is the share of interstate traffic, magnified in a complicated fashion. This percentage factor differs from state to state, but is gradually being equalized over an eight year period to about 25%. The remaining 75% of these non-traffic-sensitive (NTS) costs is split between local and intrastate toll markets using methods which differ across states.

Switching and transport costs, termed traffic sensitive costs (TS) are apportioned between intrastate and interstate jurisdictions in approximate accordance with the relative proportions of traffic in each sector. Thus, the TS cost allocation is more sensitive to changes in the composition of traffic than is the allocation of NTS costs, because the percent factor applied to the latter is divorced from changes in current period market shares.

Private line costs, including WATS access lines, are allocated

between intrastate and interstate jurisdictions on the basis of the tariff from which the service is purchased. (Recently the FCC has recognized an anomaly in its separations proceedings with respect to the allocation of the cost of private lines which carry jurisdictionally mixed traffic, and an inquiry is currently underway to determine an allocation methodology).

Representation of the Regulatory System in the Analytical Model

Qualitative Discussion

The current regulatory system, as described above, is based on an implicit assumption that the three categories of network costs, NTS, TS and private line, are additively separable. Applied to this model, the assumption would be written as:

$$8) \quad C(Z1,Z2,Z3) = C1(Z1) + C2(Z2) + C3(Z3)$$

However, this formulation precludes the possibility of economies of scope in the construction process. Therefore, the more general form will be retained.

The goal of the analysis is to adequately reflect the essential features and consequences of the current regulatory system without becoming overburdened by complicated notation. To that end, the following simplifying assumptions are made:

a) Total capitalized costs,i.e., C(Z1,Z2,Z3), are to be allocated between two sectors, local and interoffice.

b) Because of the joint nature of the construction process, "C" cannot be split apart, so that the same allocation method will apply to all capitalized costs.

c) Costs allocated to the local sector must be recovered from single line and multiline network access, and local usage demand categories.

d) Costs allocated to the interoffice sector must be recovered from toll usage and private line demand categories.

Therefore, while not conforming exactly to the current system, this regulatory model does reflect enough of the main features of the system to make the analysis of its economic effects meaningful. In particular, two critical features of the current system captured in this framework are: 1) cost allocations can be shifted between jurisdictional sectors depending on the composition of traffic, and 2) the allocation of network access capacity costs is usage based.

Mathematical Discussion

As was discussed, the current percentage allocation factor for NTS costs is, except for its gradual transition movement, a constant, or "frozen" allocator, whereas the TS allocator is dynamic, related to

current period traffic composition. The allocator used in the (modified) regulatory system contained in this model will have the form:

9) $W4 = W0 + b\ S4$

Here, W4 is the percentage of capitalized costs, C(Z1,Z2,Z3), allocated to the toll sector (X4). This percentage consists of two parts. One is a constant, W0, which is "frozen". The second component is variable, depending on the share of interoffice traffic, S4. Moreover, the interoffice traffic share is weighted by a variable, b, which can magnify or shrink the impact of traffic mix in the cost allocation process.

This formulation represents a hybrid of the pure frozen and pure dynamic allocation methods. If $b = 0$, the allocator reduces to the frozen variety, where the value of W0 could reflect an historical market share or weighted market share. If $W0 = 0$ and $b = 1$, the allocator is based on pure relative use. If, instead, $b > 1$, the allocator is an approximation for a dynamically changing SPF. By subtraction, the percentage of capital costs assigned to the local sector (X3) is,

10) $W3 = 1 - W4 = (1 - W0) - b\ S4$

Combining the preceding set of simplifying assumptions with the hybrid allocator mechanism results in the imposition of a regulatory

profitability constraint on the firm:

11) $R(X) - rM - (P4\ Z4) + L1\ F[X,Za] + L2\ H(M,a) =$

$(W4\ V4 + W3\ V3)\ C(Z)$

The difference between revenues and non-rate base costs (R&D and variable expenses) is constrained to equal the rate of return on rate base, or capitalized costs. The rate of return is a weighted average of the authorized return in the interoffice (V4) and the authorized return in the local sector (V3). The weights are the proportions of costs assigned to each of the sectors through the hybrid allocation mechanism. (The subsequent section will include R&D expenditure in the rate base).

Therefore, the objective function that the regulated exchange carrier seeks to maximize becomes:

12) $(1-U)[\ R(X) - rC(Z) - rM - (P4\ Z4) + L1\ F[X,Za] + L2\ H(M,a)\ ]$

$+ U\ (W4\ V4 + W3\ V3 - r)\ C(Z)$

where U is a regulatory multiplier which indicates how tightly or loosely the profit constraint actually binds the firm.

First Order Conditions

The following expressions are the first order profit maximizing conditions under this regulatory system:

13-a) Local usage (X3): $b S_{41}$

$$(1-U)[\ MR(3) + L1\ F(X3)\ ] \ +\ U\ [\ b\ (V3 - V4)\ ]\ N\ (1 - S3) = 0$$

The variable "N" is defined as C(Z)/(X3 + X4), or the unseparated capitalized costs per minute of use.

The second term in (13-a) is the effect of shifting costs between jurisdictions. If the share of local traffic is very high (S3 close to 1) then little additional costs can be shifted into this sector, and the size of this term becomes small. Cost shifting is also insignificant if the differential between jurisdictional authorized returns is minor, or if the cost allocation is desensitized to the traffic mix through application of an approximate frozen allocator (b close to zero).

If the authorized returns in the different sectors are close to what the firm would have earned as an unconstrained monopolist, then the regulatory profit constraint is very loose (U close to zero). Here again, the marginal effects of cost shifting would not significantly influence the firm's profit maximizing choice of local output (and, implicitly, price).

13-b) Nominal network resource "j"

$$(1-U)[\ -rC(j) + \sum F(Zj)aj\ ] + U(\ Vavg - r\ )C(j) = 0$$

Here, Vavg is shorthand for the overall weighted average return of W3 V3 + W4 V4. At the margin, the profit maximizing choice of input Zj is influenced by the differential between this weighted average return and the cost of capital, applied to the incremental stock cost of the resource.

Combining 13-a and 13-b yields an expression which more explictly shows how this regulatory system distorts the choice of local output, as compared with the decision rule that an unconstrained profit maximizing firm would follow:

13-c) $MR(3) - MC(3) =$

$$-\left\{\left[\ R\ (Vavg - r)C(j)\ /\ MP(Zj3)\ \right] + \left[\ R\ b\ (V3 - V4)\ N\ (1 - S3)\ \right]\right\}$$

where $R = U\ /\ (1-U)$.

In the absence of regulation, U, and therefore R, would be zero, hence output would be determined by the equality of marginal revenue and marginal cost. Under rate base, rate of return regulation, marginal revenue and marginal cost diverge, for two reasons. The first reason is the differential between the weighted average return and cost of capital which is earned on the marginal resources employed to produce the last unit of output. This is reflected in

the first term on the right hand side of 13-c, which has a negative sign. This makes the difference between marginal revenue and marginal cost negative, indicating the incentive for expanding output and network resource use beyond the point of pure profit maximization.

The second reason for the divergence between MR and MC is the effect of cost shifts between jurisdictions. If the local authorized return, V3, exceeds the interoffice authorized return, V4, then the second term on the right hand side of 13-c is also negative. Thus, the incentive to expand local output and widen the gap between MR and MC is re-enforced. Ofcourse, if the interoffice authorized return is higher, then the cost shifting incentive works against the incentive to utilize more of resource Zj in local production, and the net effect is an empirical issue.

Similarly, the profit maximizing level of interoffice traffic (X4) is determined by:

13-d) $MR(4) - MC(4) =$

$$- \left\{ [\, R\,(Vavg - r)C(j) \,/\, MP(Zj4) \,] - [R\, b\, (V3 - V4)\, N\, (1 - S4) \,] \right\}$$

Again, the first term on the right hand side of 13-d measures the net addition to profit earned on the marginal network capacity added to increase production of interoffice traffic. The second term shows the effect of cost shifting between the jurisdictions. As

expected, if the local authorized return exceeds the interoffice return, this second term is positive, indicating an incentive to hold back the level of interoffice production.

Because the exchange carrier's traffic mix is heavily intensive in local, the cost shifting effects would have a greater impact in affecting the level of interoffice production.

The first order condition determining aggregate R&D expenditure is:

13-e) $(1 - U)\ [\ -r + L2\ H(M)\ ] = 0$

while the efficiency parameter "aj" is determined by:

13-f) $(1 - U)\ [\ L1\ F(Zj)Zj + L2\ H(aj)\ ] = 0$

To place an intuitive interpretation on these conditions, equations 13-b and 13-e are first used to solve for the lagrangean multipliers Li and L2, respectively, which are then substituted into 13-f. This yields an expression which explains the profit maximizing level of efficiency variable aj:

13-g) $rC(j)/aj - rM(aj)/Zj = [\ R(Vavg - r)C(j) / aj\ ]$

The first term on the left hand side of 13-g is the marginal cost of increasing effective capacity "j" , through raising nominal capacity. The second term is the marginal cost of increasing

effective capacity through raising the efficiency level of a given amount of nominal capacity through planned innovational effort. In the absence of regulation these terms would be equal (see equation 7-h). In this case, however, the marginal "nominal" cost exceeds the marginal "efficiency" cost, permanently.

The interpretation is simply that the total effective capacity of resource j is being produced inefficiently through excessive reliance on augmenting nominal resources, and too little reliance, at the margin, on innovational effort.

The reason for the divergence from the ordinary profit maximizing criterion is the net differential between the (weighted average) authorized return and cost of capital earned whenever the amount of nominal network resources is expanded. The right hand side of 13-g is this marginal net differential, spread over the number of efficiency units contained in the last nominal unit added. It is a measure of the extra, regulatory created benefit the firm receives when it augments its nominal capacity. Hence, it also represents the incentive which leads to inefficient production of network capacity j.

If cumulative R&D expenditures are included in the rate base, then the constrained profit maximization problem is written as:

13-h)

$$(1 - U)[\, R(X) - rC(Z) - rM - (P4Z4) + L1\, F[X,Za] + L2\, H(M,a)\,]$$

$$+ U\, (\, W3\, V3 + W4\, V4\,)\, [\, C(Z) + M\,]$$

First order conditions for R&D expenditure, efficiency parameter aj and nominal resource Zj are, respectively:

13-i) $(-r + L2\, H(M)) + R\,(Vavg - r) = 0$

13-j) $(1 - U)\,(\, L1\, F(Zj)Zj + L2\, H(aj)\,) = 0$

13-k) $(\, -rC(j) + L1\, F(Zj)aj\,) + R\,(Vavg - r)C(j) = 0$

Solving for the lagrangean multipliers, L1 and L2, (using 13-i and 13-k), and substituting into 13-j yields an expression determining the profit maximizing amount of innovation applied to resource j:

13-l) $(Zj/r)[(1-U)r - U(Vavg-r)]\, [\, rC(j)/aj - rM(aj)/Zj\,] = 0$

As long as the regulatory multiplier is less than the ratio -- r/Vavg -- condition 13-l can be satisfied only if the second term in brackets is zero. Hence, capacity j is produced using the efficient combination of nominal resources and innovation -- if, that is, capacity j is a form of network capital included in the rate base.

The amount of innovational effort applied to non-network resources

(i.e., Zj when j = 4), is determined by:

13-m)

$$rC(j)/aj - rM(aj)/Zj = - R\ (Vavg - r)M(aj)/Zj$$

In this case, an excessive amount of innovation is applied to non-network resources, because of the net differential between authorized return and cost of capital earned on marginal R&D expenditures. Thus, in equilibrium, the marginal cost of augmenting effective capacity through innovation exceeds the marginal cost of doing so through raising the nominal amount of the non-network resource.

Summary

To summarize this section, a comparative static model of rate base rate of return regulation implies distortions in the firm's choice of profit maximizing inputs, outputs and distribution of innovational effort. It is well known that such a model leads to excessive use of the resources included in the rate base. In addition, this discussion illustrates distortions which occur in the mix of those outputs which directly determine the allocation of costs between jurisdictions with disparite authorized rates of return. If the stock of cumulative R&D expenditures is regarded as a factor of production, then its omission from the rate base leads to excessive reliance on nominal resources as the means by which

effective network capacity is augmented. Including R&D in the rate base leads to excessive application of innovational effort to non-network resources.

V) PROFIT MAXIMIZATION UNDER CEILING PRICE REGULATION

Representation of Ceiling Price Regulation in the Analytical Model

Qualitative Discussion

Recently officials intimately involved with the formation of national and state telecommunications regulatory policy have expressed interest in the application of a system of price caps to subsets of network services, as a substitute for rate base rate of return regulation. In the context of interexchange services, this notion was expressed in an address by FCC Commissioner Patrick in December, 1986. Among other things, the address discussed the possibility of establishing caps on a basic set of core services, with price flexibility provided for non-core, i.e., competitive services. Another feature of this proposal is that cost savings realized through productivity increases in the core services would be shared with consumers in the form of lower prices.

The social contract concept is essentially the same idea applied to exchange carrier services, with core services being defined as single line residential/business subscriber access. Price flexibility would be permitted on remaining network services.

This regulatory system will be represented in the model framework by specifying a price cap for single line subscriber access (X1), and complete price flexibility for the remaining network services. In order to reflect the "productivity sharing" feature mentioned previously, the level of the cap will decline with increases in the size of the efficiency parameter applicable to line side network capital. This means the price cap can be written as:

14) $CAP = CAP(a1)$,

with $d(CAP)/d(a1) < 0$

Mathematical Discussion

Under ceiling price regulation, the constrained profit maximization problem is expressed as:

15) $R(X) - rC(Z) - rM - (P4Z4) + L1\ F[X,Za] + L2\ H(M,a)$

$+ w\ [\ PX1(X) - CAP(a1)\ ]$

In this formulation, w is a lagrangean multiplier which measures the effectiveness of the price constraint. If the ceiling price is set higher than the price which would be charged by an unconstrained profit maximizing monopolist, the multiplier is zero. Presuming the price cap forces the price of single line subscriber access to a

level below the pure monopoly price, then any increase in the cap should raise the maximum attainable profit level. Hence the derivative of 15 with respect to the CAP is -w, which has to be a positive number (i.e., the change in maximum attainable profit when the price constraint is relaxed by a small amount). Therefore, the algebraic value of w is negative.

The term PX1(X) stands for the price per unit of single line access (X1). The prevailing price represents the marginal value that the market places on the last unit purchased. The marginal value of any particular output, in turn, depends not only on the quantity of that item, but also on the quantities of all related goods. For example, if local usage and single line access are compliments in society's telecommunications production function, an increase in local usage raises the marginal value of another access line. In other words, the price the market is willing to pay in order to obtain the next access line increases. Alternatively, if multiline access and single line access are considered substitutes in the aggregate social production function, then an increase in the former will lower the marginal value or benefit of the latter. The price the market is willing to pay for another single line connection will decrease as the supply of substitutes spreads.

The notation PX1(X) is just meant to express the dependence of the price of X1 upon the entire set of network services. The following derivatives reflect the interactions discussed above:

16) $dPX1/dXj > 0$ if X1 and Xj are compliments

< 0 if X1 and Xj are substitutes

$= 0$ if X1 and Xj are independent

and,

$dPX1/dX1 < 0$ "own" price effect

First Order Conditions

The profit maximizing level of single line access connections is determined by:

17-a) $MR(1) + L1\ F(X1) + w\ dPX1/dX1 = 0$

The amount of nominal resource j used in X1 production depends on:

17-b) $-rC(j) + L1\ F(Zj)aj = 0$

and combining these two conditions yields:

17-c) $MR(1) - [\ rC(j)/MP(Zj1)aj\ ] + w\ dPX1/dX1 = 0$

which is equivalent to:

17-d) $MR(1) - MC(1) + w\ dPX1/dX1 = 0$

Since the last term in 17-d has a positive sign, the condition

states that marginal revenue will be less than the marginal cost of producing X1. Hence, the firm is encouraged to produce output of single line subscriber access beyond the pure profit maximizing level.

For any output, Xi, the comparable first order condition is:

17-e) $MR(i) - MC(i) + w\, dPX1/dXi = 0$

So that if X1 and Xi are compliments, the last term in 17-e is negative. This means the production of Xi falls short of its pure profit maximizing level (MR(i) must exceed MC(i)). Alternatively, if X1 and Xi are substitutes, Xi's output will be pushed beyond its profit maximizing level.

At first, these results seem counter intuitive, because if the price of X1 is artificially lowered, one would expect that compliments to X1 would increase and substitutes would fall. In fact, the model predicts just the reverse. (NOTE - Result depends on shape of ISO-PROFIT CONTOURS)

The reason for this feasible apparent anomaly is that the firm is attempting to maximize its profit, but with a restricted set of feasible outputs that satisfy an auxilliary price constraint. Maximizing total profit while at the same time maintaining the price of X1 at an artificially low level requires a mix of outputs that is more intensive in outputs that are substitutes for X1, and less intensive in outputs that are compliments with X1.

This counter intuitive result can be illustrated by combining and rearranging the terms of 17-d and 17-e to obtain:

17-f)

$$- [MR(1) - MC(1)] / [MR(i) - MC(i)] = - (dPX1/dX1) / (dPX1/dXi)$$

The term on the right hand side of 17-f shows the marginal rate of transformation between Xi and X1 along the price constraint, i.e., the rate at which Xi can be substituted for X1, keeping the price of X1 constant. If the two outputs are substitutes, the term has a negative sign, which means X1 and Xi must move in opposite directions in order to keep PX1 the same. If the goods are compliments, the term has a positive sign, indicating that an increase in X1 must be accompanied by an increase in its compliment to keep its marginal value the same.

The term on the left hand side of 17-f shows the marginal rate of transformation between the outputs along an iso-profit contour, i.e., the rate at which Xi can be substituted for X1, holding profit constant.

Essentially, the goal of the firm is to vary its output mix along the price constraint line until the highest profit contour is reached (point of tangency between the two curves). If the goods are substitutes, the tangency must occur on the negatively sloped section of an iso-profit curve (North-East of the pure profit

maximizing point). This implies that both Xl and Xi are higher than their respective profit maximizing levels. If the goods are compliments, the tangency occurs on the positively sloped section of an iso-profit curve, (North-West of the pure profit maximizing point) implying that if Xl is higher than its profit maximizing level, Xi must be lower.

Deviations of outputs from unconstrained profit maximizing levels imply deviations in levels of nominal resources. However, conditional on the output levels and efficiency paramters, usage of nominal resources continues to be efficient. For example, the first order conditions for nominal resources j and k are:

$$17\text{-g)} \quad -rC(j) + Ll\ F(Zj)aj = 0$$

$$-rC(k) + Ll\ F(Zk)ak = 0$$

and rearranging terms yields:

$$17\text{-h)} \quad rC(j)/rC(k) = Ll\ F(Zj)aj \ /\ Ll\ F(Zk)ak$$

Thus, the ratio of marginal factor costs is equal to the marginal rate at which the nominal resources can be substituted in the multioutput production process.

For the most part, choice of efficiency parameters is also in accordance with the usual first order conditions. The exception

occurs with respect to the parameter "a1" which affects the level of the price cap. By assumption, the higher the productivity of lineside network capital, the lower the price cap. This restraint yields the following first order condition for "a1":

17-i) $rC(1)/a1 - rM(a1)/Z1 = [\ w\ d(CAP)/d(a1)\ ] / Z1$

The right hand side of 17-i represents the profit penalty imposed on the firm due to the reduction in the price cap incident to innovation in lineside network capital. This is a positive number. Hence, the left hand side of the equation shows that, at the margin, the cost of augmenting effective lineside capacity through expansion of nominal resources exceeds the cost of doing so through extra innovational effort. Too little innovation is applied because of the penalty.

Finally, using first order conditions 17-a and 17-b (output and nominal resource), condition 17-i can be rewritten as:

17-j) $MRP(Z1) - rM(a1)/Z1 = [\ w\ d(CAP)/d(a1)\ ] / Z1$

where MRP(Z1) is the weighted average marginal revenue product of effective capacity Z1, across all product lines where it is utilized. The weights are the shares of maximum nominal capacity consumed by each product line. Hence, it is a measure of the aggregate marginal productivity of an increase in "a1".

The interpretation of 17-j is that making the price cap a decreasing function of efficiency of one particular kind of network resource results in too little innovation in that sector, evidenced by the positive difference between the aggregate marginal revenue product of the innovation and its marginal R&D cost.

An alternative procedure might be to make the price cap a decreasing function of total effective capacity (Zj x aj), instead of just aj. This would eliminate the distortion in 17-i. The trade off, however, is that another distortion is created in terms of the amounts of nominal resources used. Specifically, if the price cap function took the form:

17-k) $CAP = CAP(Z1a1)$, with

$$d(CAP)/d(Z1) = [\, d(CAP)/d(Z1a1) \,]\, a1 < 0$$
$$d(CAP)/d(a1) = [\, d(CAP)/d(Z1a1) \,]\, Z1 < 0$$

the amount of resource Z1 relative to any other resource Zk would be determined by:

17-l)

$$rC(1)/rC(k) - F(Z1)a1/F(Zk)ak = -[\, w\, d(CAP)/d(Z1a1) \,]\, a1 \,/\, L1\, F(Zk)ak$$

In 17-h, the ratio of marginal factor costs of nominal resources

equaled the rate at which the inputs could be substituted in multioutput production. Here, the ratio of marginal factor costs is less than the technical marginal substitution rate, so that too little of resource Z1 is used in comparison to resource Zk.

Similar derivation performed with respect to efficiency parameters a1 and ak yield:

17-m) $F(Z1)Z1/F(Zk)Zk - H(a1)/H(ak) =$

$$- [w \, d(CAP)/d(Z1a1)] / L2 \, H(ak)$$

The first term on the left hand side of 17-m shows the marginal rate of technical substitution between efficiency paramters a1 and ak in the multioutput production process. The second term is the rate at which those paramters can be substituted within the innovational constraint (or knowledge production function). In the absence of regulation these rates of change would be equal. Here, even though the price cap now depends on total effective capacity (Z1 x a1), the continued presence of a profit penalty causes too little innovation to occur in sector 1 relative to the amount of finite knowledge applied to sector k.

Summary

To summarize this section of the discussion, a comparative static

model of ceiling price regulation was developed in which the firm maximizes profit subject to movement along a price constraint curve in output space. This requirement gave rise to the counter intuitive result that, in equilibrium, satisfaction of the price constraint on good X1 requires increased production of substitute goods and decreased production of complimentary goods. Moreover, if the price cap is specified as a decreasing function of efficiency of a particular kind of network resource, the firm has an incentive to retard the amount of innovation in that sector, relative to nominal resources, due to the profit penalty. This distortion in the sector specific application of finite technical knowledge can be eliminated if the price cap is made a decreasing function of total effective capacity of a resource. Although this procedure balances out the use of nominal and innovational efforts on a specific network segment, an incentive is created to use too little of both the nominal capacity and innovation relative to what is used on other network segments in the multioutput production process.

CONCLUSION

This paper has presented a comparative static analysis of the profit maximizing behavior of a multioutput exchange carrier under rate base rate of return regulation and ceiling price regulation. Both regulatory systems cause distortions to occur in terms of output/input composition and allocation of innovational effort. The output distortions under rate of return regulation can be eliminated by application of a frozen allocator. However, misallocation of

nominal and innovational resources still would persist, depending on the coverage of the rate base. Under ceiling price regulation, disengaging the level of the price cap from measures of network productivity would alleviate the distortion in demand for innovation in those resources utilized in the capped product line. However, the output mix would still deviate from the pure profit maximizing solution, depending on whether the capped product was a substitute or compliment to other outputs.

From the perspective of a profit maximizing firm concerned with achieving the highest possible return for its shareholders, determination of the preferred regulatory system is an empirical issue. That is, the demand and supply side relationships discussed must be incorporated into a simulation model, and quantitative comparisons made.

Of course, economic efficiency is also an important consideration, and should not be neglected. However, as new technologies erode what was traditionally the natural monopoly local franchise, the market for exchange carrier network services becomes more competitive and contestable. In such an environment, profit maximizing behavior leads to an economically efficient use of scarce resources.